



04-01-02

1632

Atty. Docket No.: 18396/1140

PATENT

#27
44-02
P. 2.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application of: Robinson, et al.
Serial No.: 09/484,629
Filed: February 18, 2000
Entitled: "Obesity Gene"

Examiner: Waitach, J.

Group Art Unit: 1632

Conf. No.: Not Yet Assigned

CERTIFICATE OF MAILING UNDER 37 CFR 1.10

I hereby certify that the paper (and any paper or fee referred to as being enclosed) is being deposited with the United States Postal Service using Express Mail to Addressee Service, under 37 C.F.R. Section 1.10, **Express Mail Label No. EL932819706US** on this date, **March 29 2002**, postage prepaid, in an envelope addressed to: Commissioner for Patents, Washington, D.C. 20231.

Kathleen M. Williams

Name of Person Mailing Paper

Signature of Person Mailing Paper

RECEIVED

Commissioner for Patents
Washington, D.C. 20231

APR 03 2002

TRANSMITTAL LETTER

TECH CENTER 1600/2900

Enclosed for filing the above-identified patent application, please find the following documents:

1. Certified Copy of International patent application number PCT/GB99/02658; and
2. Return Post Card.

The Commissioner for Patents is hereby authorized to charge any additional fees or credit any overpayment in the total fees to Deposit Account No. 16-0085, Reference 18396/1140. A duplicate of this transmittal letter is enclosed for this purpose.

Respectfully submitted,

Date: 3/29/02

Name: Kathleen M. Williams
Registration No.: 34,380
Customer No.: 29933
Palmer & Dodge LLP
111 Huntington Avenue
Boston, MA 02199-7613
Tel: 617-239-0100



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

RECEIVED

APR 03 2002

TECH CENTER 1600/2900

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation and Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the international application filed on 12 August 1999 under the Patent Cooperation Treaty at the UK Receiving Office. The application was allocated the number PCT/GB99/02658.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Date: 27 February 2002

Home

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving office use only

International Application No. **PCT/GB 99 / 02658**

International Filing Date **12 AUGUST 1999**
12 08 99

United Kingdom Patent Office
PCT International Application
Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference:
(if desired) (12 characters maximum) **P005152WO AM**

Box No. I TITLE OF INVENTION

GENE

Box No. II APPLICANT

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)

Medical Research Council
20 Park Crescent
London
W1N 4AL
United Kingdom

☐ This person is also inventor

Telephone No.

Facsimile No.

Teleprinter No.

State (i.e. country) of nationality: **United Kingdom**

State (i.e. country) of residence: **United Kingdom**

This person is applicant for the purposes of: ☐ all designated States ☒ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below.)

ROBINSON, Iain
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
United Kingdom

This person is:

☐ applicant only

☒ applicant and inventor

☐ inventor only (if this check-box is marked, do not fill in below)

State (i.e. country) of nationality: **United Kingdom**

State (i.e. country) of residence: **United Kingdom**

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

☒ Further applicant and/or (further) inventors are indicated on a continuation sheet

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:

☒ agent

☐ common representative

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)

MASCHIO, Antonio
D Young & Co
21 New Fetter Lane
London
EC4A 1DA
United Kingdom

Telephone No. **+ 44 23 8063 4816**

Facsimile No. **+44 23 8022 4262**

Teleprinter No. **477667 YOUNGS G**

☐ Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Continuation of Box No. III FURTHER APPLICANTS AND/OR (FURTHER) INVENTORS***If none of the following sub-boxes is used, this sheet is not to be included in the request.***

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

STOYE, J P
National Institute for Medical Research
The Ridgeway
Mill Hill
London NW7 1AA
United Kingdom

This person is:

- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only (if this check-box is marked, do not fill in below)

State (that is, country) of nationality:

United Kingdom

State (that is, country) of residence:

United Kingdom

This person is applicant for the purposes of:

- ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

FLAVELL, D M
Centre for Cardiovascular Genetics
The Rayne Institute
London
WC1E 6JJ
United Kingdom

This person is:

- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only (if this check-box is marked, do not fill in below)

State (that is, country) of nationality:

United Kingdom

State (that is, country) of residence:

United Kingdom

This person is applicant for the purposes of:

- ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

WELLS, S E
Department of Medicine
University of Bristol
Bristol
BS2 8HW
United Kingdom

This person is:

- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only (if this check-box is marked, do not fill in below)

State (that is, country) of nationality:

United Kingdom

State (that is, country) of residence:

United Kingdom

This person is applicant for the purposes of:

- ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: (Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)

LE TISSIER, P R
National Institute for Medical Research
The Ridgeway
Mill Hill
London
NW1AA
United Kingdom

This person is:

- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only (if this check-box is marked, do not fill in below)

State (that is, country) of nationality:

United Kingdom

State (that is, country) of residence:

United Kingdom

This person is applicant for the purposes of:

- ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on a continuation sheet

Box No. V DESIGNATION OF STATES

The following designations are hereby made under Rule 4.9(a) (mark the applicable check-boxes; at least one must be marked):

Regional Patent

- ☒ **AP** **ARIPO Patent:** GH Ghana, GM Gambia, KE Kenya, LS Lesotho, MW Malawi, SD Sudan, SZ Swaziland, UG Uganda, ZW Zimbabwe, and any other State which is a Contracting State of the Harare Protocol and of the PCT
- ☒ **EA** **Eurasian Patent:** AM Armenia, AZ Azerbaijan, BY Belarus, KG Kyrgyzstan, KZ Kazakhstan, MD Republic of Moldova, RU Russian Federation, TJ Tajikistan, TM Turkmenistan, and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT
- ☒ **EP** **European Patent:** AT Austria, BE Belgium, CH and LI Switzerland and Liechtenstein, CY Cyprus, DE Germany, DK Denmark, ES Spain, FI Finland, FR France, GB United Kingdom, GR Greece, IE Ireland, IT Italy, LU Luxembourg, MC Monaco, NL Netherlands, PT Portugal, SE Sweden, and any other State which is a Contracting State of the European Patent Convention and of the PCT
- ☒ **OA** **OAPI Patent:** BF Burkina Faso, BJ Benin, CF Central African Republic, CG Congo, CI Côte d'Ivoire, CM Cameroon, GA Gabon, GN Guinea, GW Guinea-Bissau, ML Mali, MR Mauritania, NE Niger, SN Senegal, TD Chad, TG Togo, and any other State which is a member State of OAPI and a Contracting State of the PCT (if other kind of protection or treatment desired, please specify on dotted line)

National Patent (if other kind of protection or treatment desired, specify on dotted line):

- | | |
|---|---|
| <input checked="" type="checkbox"/> AL Albania | <input checked="" type="checkbox"/> LS Lesotho |
| <input checked="" type="checkbox"/> AM Armenia | <input checked="" type="checkbox"/> LT Lithuania |
| <input checked="" type="checkbox"/> AT Austria | <input checked="" type="checkbox"/> LU Luxembourg |
| <input checked="" type="checkbox"/> AU Australia | <input checked="" type="checkbox"/> LV Latvia |
| <input checked="" type="checkbox"/> AZ Azerbaijan | <input checked="" type="checkbox"/> MD Republic of Moldova |
| <input checked="" type="checkbox"/> BA Bosnia and Herzegovina | <input checked="" type="checkbox"/> MG Madagascar |
| <input checked="" type="checkbox"/> BB Barbados | <input checked="" type="checkbox"/> MK The former Yugoslav Republic of Macedonia |
| <input checked="" type="checkbox"/> BG Bulgaria | <input checked="" type="checkbox"/> MN Mongolia |
| <input checked="" type="checkbox"/> BR Brazil | <input checked="" type="checkbox"/> MW Malawi |
| <input checked="" type="checkbox"/> BY Belarus | <input checked="" type="checkbox"/> MX Mexico |
| <input checked="" type="checkbox"/> CA Canada | <input checked="" type="checkbox"/> NO Norway |
| <input checked="" type="checkbox"/> CH AND LI Switzerland and Liechtenstein | <input checked="" type="checkbox"/> NZ New Zealand |
| <input checked="" type="checkbox"/> CN China | <input checked="" type="checkbox"/> PL Poland |
| <input checked="" type="checkbox"/> CU Cuba | <input checked="" type="checkbox"/> PT Portugal |
| <input checked="" type="checkbox"/> CZ Czech Republic | <input checked="" type="checkbox"/> RO Romania |
| <input checked="" type="checkbox"/> DE Germany | <input checked="" type="checkbox"/> RU Russian Federation |
| <input checked="" type="checkbox"/> DK Denmark | <input checked="" type="checkbox"/> SD Sudan |
| <input checked="" type="checkbox"/> EE Estonia | <input checked="" type="checkbox"/> SE Sweden |
| <input checked="" type="checkbox"/> ES Spain | <input checked="" type="checkbox"/> SG Singapore |
| <input checked="" type="checkbox"/> FI Finland | <input checked="" type="checkbox"/> SI Slovenia |
| <input checked="" type="checkbox"/> GB United Kingdom | <input checked="" type="checkbox"/> SK Slovakia |
| <input checked="" type="checkbox"/> GD Grenada | <input checked="" type="checkbox"/> SL Sierra Leone |
| <input checked="" type="checkbox"/> GE Georgia | <input checked="" type="checkbox"/> TJ Tajikistan |
| <input checked="" type="checkbox"/> GH Ghana | <input checked="" type="checkbox"/> TM Turkmenistan |
| <input checked="" type="checkbox"/> GM Gambia | <input checked="" type="checkbox"/> TR Turkey |
| <input checked="" type="checkbox"/> HR Croatia | <input checked="" type="checkbox"/> TT Trinidad and Tobago |
| <input checked="" type="checkbox"/> HU Hungary | <input checked="" type="checkbox"/> UA Ukraine |
| <input checked="" type="checkbox"/> ID Indonesia | <input checked="" type="checkbox"/> UG Uganda |
| <input checked="" type="checkbox"/> IL Israel | <input checked="" type="checkbox"/> US United States of America |
| <input checked="" type="checkbox"/> IN India | <input checked="" type="checkbox"/> UZ Uzbekistan |
| <input checked="" type="checkbox"/> IS Iceland | <input checked="" type="checkbox"/> VN Viet Nam |
| <input checked="" type="checkbox"/> JP Japan | <input checked="" type="checkbox"/> YU Yugoslavia |
| <input checked="" type="checkbox"/> KE Kenya | <input checked="" type="checkbox"/> ZW Zimbabwe |
| <input checked="" type="checkbox"/> KG Kyrgyzstan | |
| <input checked="" type="checkbox"/> KP Democratic People's Republic of Korea | |
| <input checked="" type="checkbox"/> KR Republic of Korea | |
| <input checked="" type="checkbox"/> KZ Kazakhstan | |
| <input checked="" type="checkbox"/> LC Saint Lucia | <input checked="" type="checkbox"/> AE United Arab Emirates |
| <input checked="" type="checkbox"/> LK Sri Lanka | <input checked="" type="checkbox"/> ZA South Africa |
| <input checked="" type="checkbox"/> LR Liberia | <input checked="" type="checkbox"/> CR Costa Rica X DM Dominica |

Check-boxes reserved for designating States (for the purposes of a national patent) which have become party to the PCT after the issuance of this sheet:

Precautionary Designation Statement: In addition to the designations made above, the applicant also makes under Rule 4.9(b) all other designations which would be permitted under the PCT except any designation(s) indicated in the Supplemental Box as being excluded from the scope of this statement. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit. (Confirmation of a designation consists of the filing of a notice specifying that designation and the payment of the designation and confirmation fees. Confirmation must reach the receiving Office within the 15-month time limit.)

Supplemental Box*If the Supplemental Box is not used, this sheet need not be included in the request.***Use this box in the following cases:****1. If, in any of the Boxes, the space is insufficient to furnish all the information:***in particular:*

- (i) if more than two persons are involved as applicants and/or inventors and no "continuation sheet" is available:
- (ii) if, in Box No. II or in any of the sub-boxes of Box No. III, the indication "the States indicated in the Supplemental Box" is checked:
- (iii) if, in Box No. II or in any of the sub-boxes of Box No. III, the inventor or the inventor/applicant is not inventor for the purposes of all designated States or for the purposes of the United States of America:
- (iv) if, in addition to the agent(s) indicated in Box No. IV, there are further agents:
- (v) if, in Box No. V, the name of any State (or OAPI) is accompanied by the indication "patent of addition," or "certificate of addition," or if, in Box No. V, the name of the United States of America is accompanied by an indication "Continuation" or "Continuation-in-part":
- (vi) if there are more than three earlier applications whose priority is claimed:

in such case, write "Continuation of Box No. ..." [indicate the number of the Box] and furnish the information in the same manner as required according to the captions of the Box in which the space was insufficient;

in such case, write "Continuation of Box No. III" and indicate for each additional person the same type of information as required in Box No. III. The country of the address indicated in this Box is the applicant's State (i.e. country) of residence if no State of residence is indicated below;

in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and III" (as the case may be), indicate the name of the applicant(s) involved and, next to (each) such name, State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is applicant;

in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and III" (as the case may be), indicate the name of the inventor(s) and, next to (each) such name, State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is inventor;

in such case, write "Continuation of Box No. IV and indicate for each further agent the same type of information as required in Box No. IV;

in such case, write "Continuation of Box No. V" and the name of each State involved (or OAPI), and after the name of each such State (or OAPI), the number of the parent title or parent application and the date of grant of the parent title or filing of the parent application;

in such case, write "Continuation of Box No. VI" and indicate for each additional earlier application the same type of information as required in Box No. VI.

2. If the applicant claims, in respect of any designated Office, the benefits of provisions of the national law concerning non-prejudicial disclosures or exceptions to lack of novelty:

in such case, write "Statement Concerning Non-Prejudicial Disclosures or Exceptions to Lack of Novelty" and furnish that statement below.

Continuation of Box No. IV

PURVIS, William Michael Cameron
 COTTER, Ivan John
 PILCH, Adam John Michael
 CRISP, David Norman
 ROBINSON, Nigel Alexander Julian
 HARRIS, Ian Richard
 HARDING, Charles Thomas
 TURNER, James Arthur
 MALLALIEU, Catherine Louise
 PRATT, Richard Wilson
 PRICE, Paul Anthony King
 HOLMES, Miles
 HORNER, David Richard
 NACHSHEN, Neil
 POTTER, Julian
 HAINES, Miles

Box No. VI PRIORITY CLAIM

Further priority claims are indicated in the Supplemental Box

The priority of the following earlier application(s) is hereby claimed:

Filing Date of earlier application (day/month/year)	Number of earlier application	Where earlier application is:		
		national application: country	regional application: * regional Office	international application: receiving Office
item (1) 12 Aug 1998 12/8/1998	9817566.4	UK GB ^A		
item (2) 6 May 1999 6/5/1999	9910522.3	UK GB ^A		
item (3)				

- ☒ The receiving Office is hereby requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) (only if the earlier application was filed with the Office which for the purposes of the present international application is the receiving Office) identified above as item(s): (1), (2)

* Where the earlier application is an ARIPO application, it is mandatory to indicate in the Supplemental Box at least one country party to the Paris Convention for the Protection of Industrial Property for which that earlier application was filed (Rule 4.10(b)(ii)). See Supplemental Box.

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA)
(If two or more International Searching Authorities are competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used):

ISA / EPO

Request to use results of earlier search; reference to that search (if an earlier search has been carried out by or requested from the International Searching Authority):

Date (day/month/year) _____ Number: _____ Country (or regional Office): _____

Box No. VII CHECK LIST; LANGUAGE OF FILING

This international application contains the following number of sheets:

request : 5 ✓
description (excluding sequence listing part) : 93 ✓
claims : 5 ✓
abstract : 1 ✓
drawings : 14 ✓
sequence listing part of description : 44 ✓
total number of sheets : 162 ✓

This international application is accompanied by the item(s) marked below:

- ☒ fee calculation sheet
- ☐ separate signed power of attorney
- ☐ copy of general power of attorney; reference number, if any:
- ☐ statement explaining lack of signature
- ☐ priority documents(s) identified in Box No. VI as item(s):
- ☐ translation of international application into (language):
- ☐ separate indications concerning deposited microorganism or other biological material
- ☐ nucleotide and/or amino acid sequence listing in computer readable form
- ☐ other (specify):

Figure of the drawings which should accompany the abstract:

Language of filing of the international application:

ENGLISH

Box No. IX SIGNATURE OF APPLICANT OR AGENT

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request)



MASCHIO, Antonio

1. Date of actual receipt of the purported international application:

12 AUGUST 1999 12 08 99

For receiving Office use only

3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:

4. Date of timely receipt of the required corrections under PCT Article 11(2):

5. International Searching Authority specified by the applicant:

ISA /

6. ☐ Transmittal of search copy delayed until search fee paid

2. Drawings:

☒ received:

☐ not received:

Date of receipt of the record copy by the International Bureau:

For International Bureau use only

Gene

The present invention relates to a gene which is involved in the control of obesity and fertility. In particular, the gene disclosed herein is involved in late-onset obesity in males, which is coupled with infertility. Moreover, the invention relates to animal models for late-onset obesity.

Obesity, which differs from being overweight by being characterised by an increase in the proportion of body fat present as opposed to a mere increase in body weight, is one of the major contributors to chronic disease development. Mortality in overweight males (5-15% overweight) increases to 125%, but rises up to 500% in obese males. Laboratory and epidemiological studies have also shown that mortality amongst obese males aged between 25 and 34 can increase up to twelve times. This increase in mortality is caused by the multitude of health risks associated with obesity, including cardiovascular disease, hypertension, diabetes, sleep apnoea (the abnormal ceasing of breathing during sleep), hernias, flat feet, arthritis, osteoarthritis, some cancers, varicose veins, gout, respiratory problems, gall bladder disease and liver disease. The more serious complaints include:

- Cardiovascular Disease – Obesity is an important factor in cardiovascular disease in both increasing blood cholesterol and blood pressure, and has been shown to increase the risk of disease by up to three times. Obesity also increases the work of the heart – cardiac volume, stroke volume and blood volume must all increase to cope with the increased weight. The detrimental effects of obesity on cardiovascular disease are reversible with weight loss.
- Diabetes – Excess weight also increases the chance of acquiring diabetes mellitus by threefold and increases the risk of dying from diabetes by up to eight times. The mechanism by which an increase in body fat increases the risk of diabetes is largely unknown. However, it is postulated that a slight increase in circulating serum glucose or L-leucine may increase the basal levels of insulin. This rise in insulin is then associated with resistance to insulin, caused both by decreased intracellular effects of insulin and a reduction in insulin receptors in the cell. Obesity has also been proven to

alter pancreatic function and in susceptible individuals this may lead to the development of diabetes mellitus .

- 5 • Cancer - cancer has also shown a significant association with obesity, but the mechanisms are not understood. One proposed explanation is that obesity alters hormone levels and this could influence cancer development. In males, obesity is associated with a greater risk of developing prostate and colorectal cancers.
- 10 • Gall bladder Disease – obese males show a four times increase in the risk of developing gall bladder disease.
- Endocrine Function – this is also modified by elevated fat levels. The Beta cells in the islets of Langerhans are enlarged in obese people and glucose intolerance is also frequently inhibited.
- 15 • Reproductive System – obesity in males also impairs the functioning of the reproductive system.
- Growth Hormone – obesity impairs the release of growth hormone from the pituitary gland. This problem is of particular importance in obese children whose growth may be impaired. It is fully reversible if weight is lost.
- 20

Numerous genes, gene products and their receptors have been characterised in rodent models of obesity which bear mutations associated with different forms of obesity (Bray & York, 1979; Comuzzie & Allison, 1998). Most such spontaneous mutations are recessive, and include mutations affecting leptin and its receptors in such models as ob/ob and db/db mice, Zucker fa/fa rats, Koletsky (f) rats, OLETF rats, corpulent (cp) rats and their substrains or derivatives (Zhang *et al.*, 1994; Tartaglia *et al.*, 1995; Iida *et al.*, 1996; Takaya *et al.*, 1996; Chen *et al.*, 1996; Jamal *et al.* 1997; Kahle *et al.*, 1997; Lee *et al.*, 1997; Moon & Friedman, 1997; Takiguchi *et al.*, 1998). These phenotypes are thought to result from a disruption in leptin or its receptors or in CCK-A receptors, and affect the control of food intake or energy expenditure or metabolism, and disrupt the gonadotrophic

axis in females.

There are numerous other candidate genes putatively involved in obesity, some of which have been recently been summarised by Comuzzie & Allison (1998; Table 1). These include tubby (tub), agouti, Nhl2, MCH, CRH, hypocretins or orexins, CART peptides, melanocortin-4 ligands, uncoupling proteins (UCP1-3), carboxypeptidase E, NPY, their related transcripts or homologues or their receptors (Coleman *et al.*, 1990; Miller *et al.*, 1993; Good *et al.*, 1997; Klebig *et al.*, 1995; Naggert *et al.*, 1995; Ollman *et al.*, 1995; Kleyne *et al.*, 1996; Richard, 1996; Qu *et al.*, 1996; Fan *et al.*, 1997; Huszar *et al.*, 1997; North *et al.*, 1997; Ohki-Hamazaki *et al.*, 1997; Graham *et al.*, 1997; Boss *et al.*, 1997; Vidal-Puig *et al.*, 1997; Millet *et al.*, 1997; Cool *et al.*, 1997; Kristensen *et al.*, 1998; De Lecea *et al.*, 1998; Sakurai *et al.*, 1998). These models exhibit some degree of sexual dimorphism, a slight delay in onset of obesity or a dominant pattern of inheritance, though none show all of these in combination. It is generally believed that obesity is due to the complex interaction of a number of different factors.

The study of obesity and its effects on health requires suitable animal models which can faithfully replicate the condition as seen in humans. None of the available models combines all of the symptoms of obesity. In particular, the symptoms of male pattern obesity, which include late onset, sterility and a concentration of fat around the abdomen, are not displayed by currently available models. There is therefore a need for an improved model for obesity, which displays more of the characteristics of obesity observed in human patients.

Transgenesis is a well established technique for the introduction of DNA sequences into the mammalian genome, and has been used to insert endocrine genes in several species, predominantly in mice (Palmiter *et al.*, 1982; Bucchini *et al.*, 1986; McGrane *et al.*, 1988; Ho *et al.*, 1995), but also in other species (Hammer *et al.*, 1985; Pursel *et al.*, 1989), including rats (Mullins *et al.*, 1990, Zeng *et al.*, 1994, Chareau *et al.*, 1996; Flavell *et al.*, 1996). The methods are well described (Hogan *et al.*, 1986, Chareau *et al.*, 1996) and usually involve the microinjection of cloned DNA fragments into the male pro-nuclei of eggs isolated from superovulated females. Such eggs are transferred into the oviduct of

pseudopregnant females (obtained by mating with vasectomized males) and carried to term. DNA extracted from tail clippings obtained from the progeny may be examined for the presence of specific transgene DNA. Depending on the integrity and stability of the DNA sequence, the number of integration sites and their location in the host genome, transgenes may become stably integrated in the host genome and transmitted to subsequent progeny.

If promoter and enhancer sequences are present in the transgene, the transgene may show high levels of expression in the host animals and the products may induce an endocrine phenotype that would be expected from the hormone product. For example, overexpression of human growth hormone (hGH) using a variety of heterologous non-specific promoters induces variable degrees of growth stimulation in transgenic animals (Palmiter *et al.*, 1982, 1983; Morello *et al.*, 1986; Pursel *et al.*, 1989; Shanahan *et al.*, 1989; Stewart *et al.*, 1992; Short *et al.*, 1992). However, transgene expression levels often differ between different transgenic lines made with the same insert, and the tissue specificity may vary, being highly dependent on the size of the DNA insert, the number of copies of the insert, its integrity and its integration site(s) in host DNA (Lacy *et al.*, 1983; Al-Shawi *et al.*, 1990; Huber *et al.*, 1994). Unexpected phenotypes may result, either as pathological consequences of inappropriate amounts of transgene product or its production in ectopic sites, and examples of this for hGH transgenes include glomerulosclerosis or female infertility in mice or rats (Bartke *et al.*, 1988; Brem *et al.*, 1989; Quaife *et al.*, 1989; Ninomiya *et al.*, 1994). Intentionally directed expression of a transgene to an ectopic site may also have a significant influence on the nature of the phenotype produced (Ornitz *et al.*, 1985; Baker *et al.*, 1992). This is well exemplified using hGH transgenes, since instead of an overgrowth phenotype, hGH can produce an opposite, dwarf phenotype in transgenic mice or rats when driven by a promoter that targets it to the central nervous system to induce negative feedback effects on the endogenous GH system (Hollingshead *et al.*, 1989, Banerjee *et al.*, 1994; Szabo *et al.*, 1995, Flavell *et al.*, 1996).

Other examples of endocrine transgenes include those targeting the genes for oxytocin (OT) and vasopressin (AVP) (Russo *et al.*, 1988; Habener *et al.*, 1989; Grant *et al.*,

1993a,b; Ang *et al.*, 1991, 1994; Murphy & Ho, 1995). These genes are expressed mainly in magnocellular neurones of the supraoptic (SON) and paraventricular (PVN) nuclei of the hypothalamus (Vandesande *et al.*, 1975; Young, 1992; Gainer & Wray, 1994). The expression of these hormonal peptides appears to be mutually exclusive, coexpression in the same neurone occurring only rarely (Kiyama *et al.*, 1990). A construct consisting of sequences 0.6 kb 5', 1.8 kb 3' and the entire structural gene of bovine OT directed expression to the oxytocinergic cells of the SON and the PVN, but also to the lung and Sertoli cells of the testis in transgenic mice (Ho *et al.*, 1995). The hypothalamic expression was also physiologically regulated with an increase in the abundance of the transgene transcript occurring during dehydration. The Sertoli cells are a site of peripheral expression of the endogenous OT gene in cattle but not in mice or rats. In these transgenic mice, the testicular transcripts are translated and processed (Ang *et al.*, 1994), suggesting that this construct contained regulatory elements capable of recapitulate the bovine expression pattern of OT in the mouse testis (Ang *et al.* 1991). Foo *et al.* (1994) have identified a testis-specific promoter in the rat AVP gene.

The AVP and OT genes are highly homologous in structure, and are transcribed in opposite orientations from positions closely linked in the genome within a single locus (Sausville *et al.*, 1985; Young, 1992). It is therefore possible that elements in the flanking sequence of the OT gene normally interact with those present in the nearby homologous AVP gene to regulate their mutually exclusive expression (Young *et al.*, 1990; Young, 1992). To test this theory, mice were generated bearing 1.25 kb of 5', 0.2 kb of 3' and the structural gene for bovine AVP fused, in same the same orientation as the endogenous genes, to the bovine OT transgene already described to show hypothalamic expression (Ho *et al.*, 1995). The resulting mice expressed the bovine OT transgene in the testis and lung, but lacked hypothalamic expression of this transgene and did not express the bovine AVP transgene. A further bovine OT transgene, including 3 kb of 5' sequence, the structural gene and 2.5 kb of downstream sequence was used in an attempt to overcome this repression, but no animals were generated, and it was argued that the region between 0.6 kb and 3 kb 5' of the bovine OT gene conferred a toxic effect in embryonic development which is usually repressed (Ho *et al.*, 1995). Transgenic rats have also been generated bearing fragments of the rat AVP gene with reporter genes inserted

into the third exon of the AVP gene. Transgenes containing 1.5 kb and 3kb of 5', 0.2kb of 3' and the rat AVP gene with a β -galactosidase reporter gene in the third exon also conferred expression to the testis. This was attributed to the presence of a cryptic testicular promoter within the reporter gene (Zeng *et al.*, 1994a). Clearly however, the use of small fragments of DNA containing OT or AVP sequences gives rise to unpredictable patterns of transgene expression.

One theory is that this variation may be overcome if sufficiently large DNA constructs are used, containing regions of DNA known as locus control regions (LCRs) that can direct tissue specific, position independent, copy number dependent, physiologically appropriate, expression of the transgene in the host (Grosveld *et al.*, 1987; Bonifer *et al.*, 1990; Huber *et al.*, 1994; Fujiwara *et al.*, 1997). Again this may be exemplified with hGH transgenes. An LCR region for the hGH gene has been defined by Jones *et al.* (1995). When a cosmid containing this sequence was used to generate several lines of transgenic mice, hGH was expressed in the pituitary gland in an appropriately regulated fashion, and the mice showed no overgrowth phenotype or other pathological consequences of overproduction of hGH.

Summary of the invention

A line of transgenic rats has been generated using a cosmid of rat DNA containing the genes for oxytocin (OT) and vasopressin (AVP), into which reporter genes were inserted, namely hGH (Roskam *et al.*, 1979) in the AVP gene and bovine OT mostly replacing the rat OT gene. To attempt to include LCR regions for this gene locus in our transgene constructs, larger DNA fragments containing both OT and AVP genes and larger amounts of flanking sequences were used, which were isolated from a rat cosmid library. One line of such rats, bearing at least 4 copies of this cosmid as a concatamer integrant, exhibits an unexpected and novel late onset obesity and infertility dominant phenotype that would not be predicted from the known DNA sequences present in this cosmid. This phenotype is clearly distinguishable from other obesity/infertility syndromes so far described.

Analysis of the cosmid sequences used in the transgene constructs reveals the presence of a previously unknown gene, which is responsible for the observed obesity phenotype.

5 Accordingly, in the first aspect of the present invention there is provided a 5'OT-EST polypeptide having a sequence selected from the group comprising the sequences set forth in any one of SEQ. ID. Nos. 2, 4 or 6, and sequences substantially homologous to any one of the polypeptides set forth in SEQ. ID. Nos. 2, 4 or 6.

10 In a second aspect, the invention provides a mutant of a 5'OT-EST polypeptide according to the first aspect of the invention which is capable, *in vivo*, of modulating the obesity of an animal expressing it.

15 In a third aspect, the present invention provides a nucleic acid encoding a 5'OT-EST polypeptide or mutant 5'OT-EST polypeptide according to the first aspect of the invention. Advantageously, the nucleic acid has a sequence selected from the group consisting of any one of SEQ. ID. Nos. 1, 3, 5 or 7; sequences which are hybridisable under stringent conditions with an oligonucleotide comprising 20 contiguous bases from any one of SEQ. ID. Nos. 1, 3, 5 or 7; sequences substantially homologous to any one of SEQ. ID. Nos. 1, 3, 5 or 7; and sequences complementary thereto. Stringent hybridisation conditions
20 are preferably as defined below.

In a fourth aspect, the invention provides diagnostic reagents for the detection of mutations, polymorphisms or other changes in 5'OT-EST which may predispose an individual to obesity. For example, the invention provides probes useful for amplifying 5'OT-EST
25 nucleic acids.

In a fifth aspect, the invention provides a transgenic non-human animal expressing, as a result of transgene expression, a 5'OT-EST polypeptide or mutant 5'OT-EST polypeptide according to the invention. Transgenic animals according to this aspect of the invention
30 are models for obesity in humans, and may be used for research into therapies and treatments which may be used to alleviate obesity.

Brief Description of the Drawings

Figure 1 shows a partial restriction map of the rat AVP/OT locus, from cosmid cVO1, cVO2 and cVO3. Restriction sites for the enzymes listed are shown as a vertical marks.

- 5 The known sequence of the rat AVP and OT genes is indicated by single dashed lines, and the sequence determined and disclosed herein of the rat 5'OT-EST gene is indicated by the double dashed line. Scale is approximate.

- 10 Figure 2 is a diagrammatic representation of the subcloning steps leading to the insertion of the hGH reporter gene into the 5' untranslated region of the rat vasopressin gene. The final subclone shows the Cla 1 to Xho 1 fragment which was inserted into the construct used to make transgenic lines.

- 15 Figure 3 is a diagrammatic representation of the subcloning steps required for the production of the rat-bovine hybrid gene which was inserted into the final construct. For simplicity, this is shown in reversed orientation compared to its orientation in the construct.

- 20 Figure 4 shows the extent of the rat AVP/OT locus present in the cosmid cVO1, 2 and 3. These clones span a total of 44kb, including 8kb 5' of rAVP and 24kb 5' of rOT. The structure of the final cosmid construct CVO14 is illustrated and some restriction sites indicated.

- 25 Figure 5 shows a point mutation in the cVO14 construct in a conserved region 5' to the OT gene. A conserved G residue is substituted with an A residue in the construct.

Figure 6 is an alignment of the sequences of 5'OT-EST from rat, human and mouse sources.

- 30 Figure 7 is a comparison of the body weights of transgenic and non-transgenic rats.

Figure 8 is a comparison of the body weights of transgenic and non-transgenic male and female rats.

Figure 9 is a comparison of the measurements (in mm) of the pelvis (b) and of the body length (a) of 20 and 52 week male transgenic and non-transgenic rats (mean \pm sem, ***= $p < 0.001$, $n=6-7$ per group).

Figure 10 illustrates the increased body weight/body length ratio of transgenic rats compared with non-transgenic rats.

10

Figure 11 shows the weights of the peri-renal and testicular fat pads in JP17 male transgenic and non-transgenic animals at different ages. (*= $p < 0.05$; **= $p < 0.01$, $n=6-7$ per group).

Figure 12 shows the levels of plasma insulin, glucose, cholesterol, triglycerides, leptin and corticosterone in terminal blood samples from transgenic and non-transgenic rats. Values shown are mean of each group \pm SEM ($n=6$ for transgenic groups; $n=4$ for non-transgenic groups) (* Significantly different ($p < 0.05$) from sex matched non transgenic group).

20

Figure 13 shows the changes in body weight, leptin levels and food intake associated with young (100 day old) rats fed on normal fat (4%) or high fat (30%) diets. Results are shown for SLOB, non-transgenic and dwarf rats, fed either 4% fat diet (clear bars) or a 30% fat diet (stippled bars) over a 27-day period (* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$, high vs. low fat diet: ## = $p < 0.01$; ### = $p < 0.001$, SLOB vs. non-transgenic rats).

Figure 14 shows the changes in body weight associated with ovariectomy in transgenic rats and non-transgenic littermates. ■ - ovariectomised SLOB rat; □ - ovariectomised wild-type rat; ● - sham ovariectomised SLOB rat; ○ - sham ovariectomised wild-type rat.

30

Detailed Description of the Invention

Definitions

- 5 As referred to herein, "5'OT-EST" is the polypeptide represented in SEQ. ID. Nos. 2, 4 or 6 (rat, human and mouse respectively). Preferably, it is the human sequence. However, the term also covers alternative peptides homologous to 5'OT-EST, such as polypeptides derived from other species, including other mammalian species.
- 10 "Mutants" of 5'OT-EST include polypeptides which differ only in minor, insignificant ways from wild-type 5'OT-EST, for example polypeptides having conservative amino acid replacements or additions or deletions. Preferred, however, are mutants which are able to confer, on animals expressing them, an obese phenotype as defined herein. An example of such a mutant is the 5'OT-EST - xdel polypeptide set forth in SEQ. ID. No. 8.
- 15 Further mutants may be obtained as described herein, and defined according to their functional effects in transgenic animals or host cells.

- "Substantially homologous", whether applied to polypeptide or nucleotide sequences, is as defined herein with reference to homology screening. It may be interpreted as referring
- 20 either to sequence alignment and direct comparison, or to homology as defined by BLAST homology searching as defined herein.

- A "transgenic animal" is an animal whose genome has been functionally altered by genetic manipulation. In the context of the present invention, this includes animals
- 25 bearing and expressing a 5'OT-EST or mutant 5'OT-EST transgene, animals from which 5'OT-EST sequences have been deleted or in which they have been modified, and animals which are transiently transformed to express a (mutant) 5'OT-EST transgene such as by transformation with viral sequences.

- 30 "Transformation" refers to the functional insertion of a gene by nucleic acid transfer, or the functional deletion of a gene, in a cell or organism. The term thus includes transfection, transduction and any other techniques useful for transferring nucleic acids

into cells or organisms. Cells transformed according to the invention express a novel genotype as a result of the transformation.

For the avoidance of doubt, unless otherwise required by the specific context, reference
5 herein to an entity in the singular includes the plural thereof. Thus, the expressions "a gene" and "one or more genes" are equivalent.

Moreover, unless otherwise required by context, references to 5'OT-EST (*5'OT-EST*)
10 preferably include mutants of 5'OT-EST (*5'OT-EST*).

A "cosmid" is a bacteriophage-based vector as commonly known in the art.

References herein to "obesity" and obese animals are preferably references to the SLOB
phenotype observed in SLOB rats according to the invention, characterised in being *inter*
15 *alia* male-specific, late onset, with fat deposition concentrated in the abdominal area and associated with sterility.

Description of Preferred Embodiments

20 A cosmid (cVO14) of rat DNA containing the rat vasopressin (AVP) and rat oxytocin (OT) genes (Ivell & Richter, 1984) was constructed, and DNA reporter sequences inserted therein using standard methods (Sambrook *et al.*, 1989) as outlined in Examples 1 & 2 below. Microinjection of the cVO14 DNA insert into fertilised rat eggs and their transfer
25 into pseudopregnant recipients resulted in production of viable offspring. Unexpectedly, the male founder rat with 4-5 copies of cVO14 (JP17) showed a dominant phenotype of severe late-onset visceral obesity. This form of obesity shows (i) a very late onset, (ii) a highly selective visceral distribution of fat developing on a normal rodent diet, without hyperphagia, (iii) an effect greatly preponderant in males, (iv) a predisposition to excessive dietary-fat induced obesity at an early age, before the phenotype becomes
30 apparent on a normal diet, and (v) a dominant pattern of inheritance. Moreover, male transgenics show severe infertility in males, whilst females are fertile. Rats bearing this transgene have been termed SLOB rats (for Severe Late-onset OBesity). The symptoms

of obesity observed in SLOB rats all occur in several forms of human obesity, including that associated with human syndrome-X (Reaven *et al.*, 1988) for which a late-onset increase in abdominally distributed fat, affecting males much more severely than females (Gray *et al.* 1997) may be mimicked in the SLOB rat. Obvious causes, such as leptin deficiency or insulin resistance or overt Type 1 or 2 diabetes may be excluded.

Although the SLOB phenotype is preponderant in males, it may be markedly exacerbated in females by ovariectomy.

Mapping and analysis of cosmid DNA used to generate cVO14, revealed a putative gene, 5' of the OT locus. A fragment of this DNA was subcloned and sequenced. Analysis of this region of rat DNA enabled us to determine the location, orientation, partial exon structure and predicted protein product of a novel gene lying 5' of the OT gene in rat DNA. Further sequencing and analysis elucidated the structure of this gene, and provided additional sequence information for the cosmid DNA surrounding the known sequence of the OT and AVP genes. The novel rat gene is termed herein *5'OT-EST*, which encodes the *5'OT-EST* polypeptide. The genomic sequence of *5'OT-EST* is given in SEQ. ID. No. 16.

A search of DNA and protein databases revealed no significant match to any known gene, but recognised partial matches to DNA sequences homologous to *5'OT-EST* in expressed sequence tag (EST) databases from rat, mouse and human DNA sources. These represent partial products of the rat gene, and of genes homologous to this novel rat gene, in mouse and human DNA. The predicted structures of four exons, termed w, x, y, z, and predicted protein sequences are highly conserved between these species. A partial match was noted to a human genomic DNA sequence alluded to, but not disclosed in White *et al.* PNAS 95:305-309 (1998), but deposited by them in Genbank (Accession no:AF036329) as a putative genomic fragment containing the human *GnRH-II* gene. The relationship between human *5'OT-EST* and human *GnRH-II* as described by White *et al.* is confirmed by the present work; however, there does not appear to be any such relationship in rats or mice. Homologous rat *GnRH-II* sequences cannot be recognised by sequence analysis in cVO14, which contains more than 10kb of rat DNA flanking *5'OT-EST*. Neither can any

homologous mouse *GnRH-II* sequence be identified by hybridisation or PCR studies in multiple mouse genomic clones which contain 5'OT-EST and at least 50kb of flanking DNA. Thus, it appears highly unlikely that the *GnRH-II* sequence corresponds functionally with 5'OT-EST.

5

In rats, 5'OT-EST lies about 10kb downstream of the 3' exon of the protein tyrosine phosphatase receptor alpha (*Ptptra*) gene, and the intervening 10kb show no homology with *GnRH-II*. Additionally, mouse BAC clones containing 5'OT-EST show no homology to *GnRH-II*. Thus, *GnRH-II* sequences are not adjacent to 5'OT-EST in rats or mice, and neither *GnRH-II* nor *Ptptra* is present in the cosmid used to generate SLOB rats. Complete sequencing of the cosmid reveals no other novel genes.

10

Based on physical linkage to *Ptptra*, *Avp* and *Oxt*, 5'OT-EST maps to the distal region of mouse chromosome 2, 7.32 cM from the centromere. *Ptptra* has itself been implicated in the control of insulin sensitivity, and both 5'OT-EST and *Ptptra* lie within 0.21 cM of *mg*, another gene implicated in the suppression of obesity. In mouse, all three genes map to the same region as the mouse obesity locus *Mob5* (Encyclopaedia of the Mouse genome VII: Mouse Chromosome 2; (1998) Peters *et al.*, Mamm. genome 8 Spec No:S27-49). It is likely therefore that 5'OT-EST contributes to the trait observed at this locus in mice.

15

20

Accordingly, the present invention provides 5'OT-EST polypeptide. 5'OT-EST according to the present invention may be mouse, rat or human 5'OT-EST, as well as variants of 5'OT-EST derivable from other species or by natural or artificial mutation of a 5'OT-EST gene.

25

The variant provided by the present invention includes splice variants encoded by mRNA generated by alternative splicing of a primary transcript, amino acid mutants, glycosylation variants and other covalent derivatives of 5'OT-EST which retain the physiological and/or physical properties thereof. Exemplary derivatives include molecules wherein 5'OT-EST is covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid. Such a moiety may be a detectable moiety such as an enzyme or a

30

radioisotope. Further included are naturally occurring variants of 5'OT-EST found within a particular species, preferably a mammal. Such a variant may be encoded by a related gene of the same gene family, by an allelic variant of a particular gene, or represent an alternative splicing variant of 5'OT-EST.

5

Variants which retain common structural features can be fragments of 5'OT-EST. Fragments of 5'OT-EST comprise smaller polypeptides derived from therefrom. Preferably, smaller polypeptides derived from 5'OT-EST according to the invention define a single feature which is characteristic of 5'OT-EST as described in the present application.

10

Derivatives of 5'OT-EST also comprise mutants thereof, which may contain amino acid deletions, additions or substitutions. Thus, conservative amino acid substitutions may be made substantially without altering the nature of 5'OT-EST. Deletions and substitutions may moreover be made to the fragments of 5'OT-EST comprised by the invention.

15

Mutants of 5'OT-EST according to the present invention may possess properties different from those of naturally occurring 5'OT-EST. In particular, 5'OT-EST mutants may modulate the expression of native 5'OT-EST.

20

5'OT-EST mutants may be produced from a nucleic acid encoding 5'OT-EST which has been subjected to in vitro mutagenesis resulting e.g. in an addition, exchange and/or deletion of one or more amino acids. For example, substitutional, deletional or insertional variants of 5'OT-EST can be prepared by recombinant methods and screened for immuno-crossreactivity with the native forms of 5'OT-EST.

25

Preferably, 5'OT-EST according to the present invention has the sequence of SEQ. ID. No. 2 (rat), SEQ. ID. No. 4 (human) or SEQ. ID. No. 6 (mouse). Mutants possessing desired properties may be generated from these sequences, or isolated from natural sources, by a variety of techniques which assess the biological function of the 5'OT-

30

EST mutant. For example, nucleic acids encoding 5'OT-EST mutants may be used to generate transgenic animals and these animals assessed for indications of an obesity phenotype.

- 5 For example, the effects of mutant transgenes may be assessed by carcass analysis, measurement of growth, body weight, body fat distribution, as well as other measures of analytes in body fluids or tissues relevant to obesity in transgenic animals (Mathe, 1995; Shillabeer, 1992). These include, but are not limited to, cholesterol, triglycerides, fatty acids, lipoproteins, and other dietary constituents or metabolites, as well as metabolic hormones, such as leptin, insulin, glucagon, catecholamines or glucocorticoids. Other relevant parameters include cardiovascular measures (Reaven, 1988, Gray & Yudkin, 1997). These may include measures of systolic or diastolic blood pressure, cardiac output, or vascular resistance, together with morphological changes to organ systems known to be affected by cardiovascular or obesity disorders, such as heart, major or minor blood vessels, their muscle or endothelial layers, and their elasticity or fragility. See for example McNamee *et al.* (1994).

- Similarly, parameters related to the infertility phenotype that may be measured, include, but are not limited to, testicular weight, volume, development, spermatogenesis, sperm number, motility or ability to fertilise oocytes. They may also include measures of testicular fluid production and constituents, as well as products of other accessory organs including seminal vesicles or prostate, as well as hormones, receptors, and proteins important in male sexual function, such as testosterone, LH, FSH, inhibin or activin. Other responses that may be affected include energy expenditure, physical activity, ingestive behaviour, excretory behaviour, or reproductive behaviour, or the organs, hormones or receptors commonly recognised to be associated with these physiological systems, their metabolism or morphological structure.

- 5'OT-EST as disclosed herein is a polypeptide composed of four exons, termed w, x, y and z (see SEQ. ID. No. 16). Advantageously, mutants of 5'OT-EST are mutated in, or preferably lack all or part of, the sequences encoded by one or more exons of 5'-OT-EST. Preferably, mutants of 5'OT-EST lack, or are mutated in, all or part of the sequences

encoded by exons x, y and z of 5'-OT-EST.

Preferably, the sequences encoded by exons x, y and z are deleted and those encoded by exon w partially deleted. Most preferably, the mutant is 5'OT-EST - xdel as described
5 herein, for example in SEQ. ID. No. 8.

The fragments, mutants and other derivatives of 5'OT-EST preferably retain substantial homology with 5'OT-EST. As used herein, "homology" means that the two entities share sufficient characteristics for the skilled person to determine that they are similar
10 in origin and function. Preferably, homology is used to refer to sequence identity. Thus, the derivatives of 5'OT-EST preferably retain substantial sequence identity with 5'OT-EST.

"Substantial homology", where homology indicates sequence identity, means more than
15 40% sequence identity, preferably more than 45% sequence identity and most preferably a sequence identity of 50% or more, as judged by direct best-fit sequence alignment and comparison.

Sequence homology (or identity) may moreover be determined using any suitable
20 homology algorithm, using for example default parameters. Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail at http://www.ncbi.nih.gov/BLAST/blast_help.html, which is incorporated herein by reference. The search parameters are defined as follows, and are advantageously set to the defined default parameters.

25 Advantageously, "substantial homology" when assessed by BLAST equates to sequences which match with an EXPECT value of at least about 7, preferably at least about 9 and most preferably 10 or more. The default threshold for EXPECT in BLAST searching is usually 10.

30 BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs

ascribe significance to their findings using the statistical methods of Karlin and Altschul (see http://www.ncbi.nih.gov/BLAST/blast_help.html) with a few enhancements. The BLAST programs were tailored for sequence similarity searching, for example to identify homologues to a query sequence. The programs are not generally useful for motif-style searching. For a discussion of basic issues in similarity searching of sequence databases, see Altschul *et al.* (1994).

The five BLAST programs available at <http://www.ncbi.nlm.nih.gov> perform the following tasks:

10

blastp compares an amino acid query sequence against a protein sequence database;

blastn compares a nucleotide query sequence against a nucleotide sequence database;

15 **blastx** compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database;

tblastn compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).

20

tblastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

BLAST uses the following search parameters:

25

HISTOGRAM Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual).

DESCRIPTIONS Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page). See also **EXPECT** and **CUTOFF**.

30

ALIGNMENTS Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting
5 (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual).

EXPECT The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found
10 merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See parameter E in the BLAST Manual).

15 CUTOFF Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher
20 CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT.

MATRIX Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and
25 TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response.

STRAND Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence.

- 5 FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of
10 Tatusov and Lipman (see <http://www.ncbi.nlm.nih.gov>). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

15

Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNNN") and the letter "X" in protein sequences (e.g., "XXXXXXXXXX").

- 20 Filtering is only applied to the query sequence (or its translation products), not to database sequences. Default filtering is DUST for BLASTN, SEG for other programs.

- It is not unusual for nothing at all to be masked by SEG, XNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an
25 effect. Furthermore, in some cases, sequences are masked in their entirety, indicating that the statistical significance of any matches reported against the unfiltered query sequence should be suspect.

- NCBI-gi Causes NCBI gi identifiers to be shown in the output, in addition to the
30 accession and/or locus name.

Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm provided at <http://www.ncbi.nlm.nih.gov/BLAST>.

Conventional BLAST searches of the publically available databases do not reveal any
5 homology of the predicted protein product of *5'OT-EST* to any known protein.
However, application of a more sophisticated search algorithm, as described in Taylor
et al., 1998, identifies structural similarities to apolipoprotein E (ApoE) in its alpha-
helical domains, but without any apparent LDL-receptor domain. Since ApoE is
centrally involved in lipid metabolism and transport, a role for *5'OT-EST* in cellular
10 lipid handling is suggested.

Accordingly, the invention provides a method for identifying a candidate compound
capable of influencing lipid transport, comprising the steps of:

- 15 a) contacting *5'OT-EST* polypeptide with a candidate compound or compounds and
determining which candidate compound or compounds is capable of interacting with
5'OT-EST;
- b) optionally, testing candidate compounds which interact with *5'OT-EST* in a
transgenic animal according to the invention.

20

According to a further aspect of the present invention, there is provided a nucleic acid
encoding *5'OT-EST* or a mutant thereof. In addition to being useful for the production of
recombinant *5'OT-EST* protein, these nucleic acids are also useful as probes, thus readily
enabling those skilled in the art to identify and/or isolate nucleic acid encoding *5'OT-EST*
25 and/or mutant *5'OT-EST*. The nucleic acid may be unlabelled or labelled with a
detectable moiety. Furthermore, nucleic acid according to the invention is useful e.g. in a
method determining the presence of *5'OT-EST*-specific nucleic acid, said method
comprising hybridising the DNA (or RNA) encoding *5'OT-EST* (or its complement) to
test sample nucleic acid and determining the presence of *5'OT-EST*. In another aspect, the
30 invention provides a nucleic acid sequence that is complementary to, or hybridises under
stringent conditions to, a nucleic acid sequence encoding *5'OT-EST*.

The invention also provides a method for amplifying a nucleic acid test sample comprising priming a nucleic acid polymerase (chain) reaction with nucleic acid corresponding to 5'OT-EST, including the untranslated regions (or its complement).

- 5 In still another aspect of the invention, the nucleic acid is DNA and further comprises a replicable vector comprising the nucleic acid encoding 5'OT-EST operably linked to control sequences recognised by a host transformed by the vector. Furthermore the invention provides host cells transformed with such a vector and a method of using a nucleic acid encoding 5'OT-EST to effect the production of 5'OT-EST, comprising
10 expressing 5'OT-EST nucleic acid in a culture of the transformed host cells and, if desired, recovering 5'OT-EST from the host cell culture.

- Isolated 5'OT-EST nucleic acid includes nucleic acid that is free from at least one contaminant nucleic acid with which it is ordinarily associated in the natural source of
15 5'OT-EST nucleic acid or in crude nucleic acid preparations, such as DNA libraries and the like. Isolated nucleic acid thus is present in other than in the form or setting in which it is found in nature. However, isolated 5'OT-EST encoding nucleic acid includes 5'OT-EST nucleic acid in ordinarily 5'OT-EST-expressing cells where the nucleic acid is in a chromosomal location different from that of natural cells or is otherwise flanked by a
20 different DNA sequence than that found in nature.

- In accordance with the present invention, there are provided isolated nucleic acids, e.g. DNAs or RNAs, encoding 5'OT-EST, particularly mammalian 5'OT-EST, e.g. human 5'OT-EST, or fragments thereof. In particular, the invention provides a DNA molecule
25 encoding 5'OT-EST, or a fragment thereof. By definition, such a DNA comprises a coding single stranded DNA, a double stranded DNA of said coding DNA and complementary DNA thereto, or this complementary (single stranded) DNA itself. An exemplary nucleic acid encoding 5'OT-EST is represented in SEQ ID Nos. 1, 3 and/or 5.

- 30 The preferred sequence encoding 5'OT-EST is that having substantially the same nucleotide sequence as the coding sequences in SEQ ID Nos. 1, 3 and/or 5, with the nucleic acid having the same sequence as the coding sequence in SEQ ID Nos. 1, 3 and/or

5 being most preferred. As used herein, nucleotide sequences which are substantially the same share at least about 90% identity. However, in the case of splice variants having e.g. an additional exon sequence homology may be lower. Homology is determined as described above.

5

The invention moreover provides nucleic acids encoding 5'OT-EST, comprising the gene 5'OT-EST or variants thereof as defined herein. The nucleic acids of the invention, whether used as probes or otherwise, are preferably substantially homologous to the sequence of 5'OT-EST as shown in SEQ ID Nos. 1, 3 and/or 5. The terms "substantially" and "homologous" are used as hereinbefore defined with reference to the 5'OT-EST polypeptide.

10

Preferably, nucleic acids according to the invention are fragments of the 5'OT-EST-sequence, or derivatives thereof as hereinbefore defined in relation to polypeptides.

15

Fragments of the nucleic acid sequence of a few nucleotides in length, preferably 5 to 150 nucleotides in length, are especially useful as probes.

Exemplary nucleic acids can alternatively be characterised as those nucleotide sequences which encode a 5'OT-EST protein, or which correspond to untranslated regions of 5'OT-EST, and hybridise to the DNA sequences set forth SEQ ID Nos. 2, 4 and/or 6, or a selected fragment of said DNA sequence. Preferred are such sequences encoding 5'OT-EST which hybridise under high-stringency conditions to the sequence of SEQ ID Nos. 1, 3 and/or 5.

20

Stringency of hybridisation refers to conditions under which polynucleic acids hybrids are stable. Such conditions are evident to those of ordinary skill in the field. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrid which decreases approximately 1 to 1.5°C with every 1% decrease in sequence homology. In general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridisation reaction is performed under conditions of higher stringency, followed by washes of varying stringency.

25

30

As used herein, high stringency refers to conditions that permit hybridisation of only those nucleic acid sequences that form stable hybrids in 1 M Na⁺ at 65-68 °C. High stringency conditions can be provided, for example, by hybridisation in an aqueous solution containing 6x SSC, 5x Denhardt's, 1 % SDS (sodium dodecyl sulphate), 0.1 Na⁺ pyrophosphate and 0.1 mg/ml denatured salmon sperm DNA as non specific competitor.
5 Following hybridisation, high stringency washing may be done in several steps, with a final wash (about 30 min) at the hybridisation temperature in 0.2 - 0.1x SSC, 0.1 % SDS.

Moderate stringency refers to conditions equivalent to hybridisation in the above described solution but at about 60-62 °C. In that case the final wash is performed at the
10 hybridisation temperature in 1x SSC, 0.1 % SDS.

Low stringency refers to conditions equivalent to hybridisation in the above described solution at about 50-52 °C. In that case, the final wash is performed at the hybridisation
15 temperature in 2x SSC, 0.1 % SDS.

It is understood that these conditions may be adapted and duplicated using a variety of buffers, e.g. formamide-based buffers, and temperatures. Denhardt's solution and SSC are well known to those of skill in the art as are other suitable hybridisation buffers (see, e.g.
20 Sambrook, *et al.*, eds. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York or Ausubel, *et al.*, eds. (1990) *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.). Optimal hybridisation conditions have to be determined empirically, as the length and the GC content of the probe also play a role.

25 Advantageously, the invention moreover provides nucleic acid sequence which are capable of hybridising, under stringent conditions, to a fragment of SEQ. ID. Nos. 1, 3, 5 or 7. Preferably, the fragment is between 15 and 50 bases in length. Advantageously, it is about 25 bases in length, preferably about 20 bases in length. For differentiating between mutant and wild type 5'*OT-EST* by PCR reactions, 20mers are the preferred size, whilst
30 for use as probes in, for example, Southern hybridisation, the use of 40mers is preferred. Riboprobes may be designed to be substantially any length, up to and including the entire length of the largest specific cDNA sequence.

Specifically included, moreover, are sequences complementary to the foregoing sequences.

5 Given the guidance provided herein, the nucleic acids of the invention are obtainable according to methods well known in the art. For example, a DNA of the invention is obtainable by chemical synthesis, using polymerase chain reaction (PCR) or by screening a genomic library or a suitable cDNA library prepared from a source believed to possess *5'OT-EST* and to express it at a detectable level.

10

Chemical methods for synthesis of a nucleic acid of interest are known in the art and include triester, phosphite, phosphoramidite and H-phosphonate methods, PCR and other autoprimer methods as well as oligonucleotide synthesis on solid supports. These methods may be used if the entire nucleic acid sequence of the nucleic acid is known, or the
15 sequence of the nucleic acid complementary to the coding strand is available. Alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid sequences using known and preferred coding residues for each amino acid residue.

An alternative means to isolate the gene encoding *5'OT-EST* is to use PCR technology as
20 described e.g. in section 14 of Sambrook *et al.*, 1989. This method requires the use of oligonucleotide probes that will hybridise to *5'OT-EST* nucleic acid. Strategies for selection of oligonucleotides are described below.

Libraries are screened with probes or analytical tools designed to identify the gene of
25 interest or the protein encoded by it. For cDNA expression libraries suitable means include monoclonal or polyclonal antibodies that recognise and specifically bind to *5'OT-EST*; oligonucleotides of about 20 to 80 bases in length that encode known or suspected *5'OT-EST* cDNA from the same or different species; and/or complementary or homologous cDNAs or fragments thereof that encode the same or a hybridising gene.
30 Appropriate probes for screening genomic DNA libraries include, but are not limited to oligonucleotides, cDNAs or fragments thereof that encode the same or hybridising DNA; and/or homologous genomic DNAs or fragments thereof.

A nucleic acid encoding 5'OT-EST may be isolated by screening suitable cDNA or genomic libraries under suitable hybridisation conditions with a probe, i.e. a nucleic acid disclosed herein including oligonucleotides derivable from the sequences set forth in SEQ ID Nos. 1, 3 and/or 5. Suitable libraries are commercially available or can be prepared e.g. from cell lines, tissue samples, and the like.

As used herein, a probe is e.g. a single-stranded DNA or RNA that has a sequence of nucleotides that includes between 10 and 50, preferably between 15 and 30 and most preferably at least about 20 contiguous bases that are the same as (or the complement of) an equivalent or greater number of contiguous bases set forth in SEQ ID Nos. 1, 3 and/or 5. The nucleic acid sequences selected as probes should be of sufficient length and sufficiently unambiguous so that false positive results are minimised. The nucleotide sequences are usually based on conserved or highly homologous nucleotide sequences or regions of 5'OT-EST. The nucleic acids used as probes may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular importance where a library is screened from a species in which preferential codon usage in that species is not known.

Preferred regions from which to construct probes include 5' and/or 3' coding sequences, sequences predicted to encode ligand binding sites, and the like. For example, either the full-length cDNA clone disclosed herein or fragments thereof can be used as probes. Preferably, nucleic acid probes of the invention are labelled with suitable label means for ready detection upon hybridisation. For example, a suitable label means is a radiolabel. The preferred method of labelling a DNA fragment is by incorporating α - 32 P dATP with the Klenow fragment of DNA polymerase in a random priming reaction, as is well known in the art. Oligonucleotides are usually end-labelled with γ - 32 P-labelled ATP and polynucleotide kinase. However, other methods (e.g. non-radioactive) may also be used to label the fragment or oligonucleotide, including e.g. enzyme labelling, fluorescent labelling with suitable fluorophores and biotinylation.

Probes for cloning and amplifying *5'OT-EST*, especially human *5'OT-EST*, may be deduced from the sequence thereof provided herein. Preferred probes may be selected from the following:

5	1U	GGACAGCCCGAAGGACTACAGGT	SEQ. ID. No. 18
	1L	CGAAGAACTCCGCAGGGTCC	SEQ. ID. No. 19
	2U	AAGACCCGCCACGACCCG	SEQ. ID. No. 20
	2L	GAATCAGCACCTCTCCGCC	SEQ. ID. No. 21
10			
	3U	TGCGGAGTTCTTCGTGCTGATGGAG	SEQ. ID. No. 22
	3L	GGTGCTCGGCGGCGTCCTTC	SEQ. ID. No. 23
	4U	GAGTGGCGGAGAGGGTGCTGA	SEQ. ID. No. 24
15	4L	GGCCGAGGCTGAGCGGGG	SEQ. ID. No. 25
	5U	CTGAAGGACGCCGCCGAGCA	SEQ. ID. No. 26
	5L	CTCCAACGCCTGCCGCTGC	SEQ. ID. No. 27
20	6U	GCAGGAGGAGCGGGAGCAGGA	SEQ. ID. No. 28
	6L	TCCAGTGCCCCGCAAGCCG	SEQ. ID. No. 29

Probes according to the invention are suitable for use as diagnostic reagents to amplify *5'OT-EST* and thereby enable the analysis of the nucleic acid for the presence of mutations, polymorphisms or other changes which could render an individual susceptible to obesity.

After screening the library, e.g. with a portion of DNA including substantially the entire *5'OT-EST*-encoding sequence or a suitable oligonucleotide based on a portion of said DNA, positive clones are identified by detecting a hybridisation signal; the identified clones are characterised by restriction enzyme mapping and/or DNA sequence analysis,

and then examined, e.g. by comparison with the sequences set forth herein, to ascertain whether they include DNA encoding a complete 5'OT-EST (i.e., if they include translation initiation and termination codons). If the selected clones are incomplete, they may be used to rescreen the same or a different library to obtain overlapping clones. If the library is genomic, then the overlapping clones may include exons and introns. If the library is a cDNA library, then the overlapping clones will include an open reading frame. In both instances, complete clones may be identified by comparison with the DNAs and deduced amino acid sequences provided herein.

- 10 In order to detect any abnormality of endogenous 5'OT-EST, genetic screening may be carried out using the nucleotide sequences of the invention as hybridisation probes or as PCR primers, using which genomic nucleic acid may be amplified, and subsequently sequenced. Also, based on the nucleic acid sequences provided herein antisense-type therapeutic agents may be designed.

- 15 It is envisaged that the nucleic acid of the invention can be readily modified by nucleotide substitution, nucleotide deletion, nucleotide insertion or inversion of a nucleotide stretch, and any combination thereof. Such mutants can be used e.g. to produce a 5'OT-EST mutant that has an amino acid sequence differing from the 5'OT-EST sequences as found in nature. Mutagenesis may be predetermined (site-specific) or random. A mutation which is not a silent mutation must not place sequences out of reading frames and preferably will not create complementary regions that could hybridise to produce secondary mRNA structure such as loops or hairpins.

- 25 The invention accordingly specifically includes nucleic acids encoding mutants of 5'OT-EST, as defined above. Such nucleic acids may be used for all the purposes identified above in relation to wild-type 5'OT-EST nucleic acids. Particularly preferred are nucleic acids encoding 5'OT-EST - *xdel*, which preferably have the sequence
ATGTTGCGGGCTTTGAACCGCCTGGCCGCGCGGCCCGGGGGCCAGCCCCAACCCTGCTC
30 CTTCTGCCCCGTGCGCGGCCACGGCCCCGCTCATTCTCGGCTCCTTTTTCCTCGCAGGAT
AGC (see SEQ. ID. No. 7), or an equivalent sequence which encodes the same polypeptide having regard to the degeneracy of the nucleic acid code, or a sequence

substantially homologous thereto or complementary thereto. In 5'-OT-EST - xdel exon x is deleted, and exons y and z are out of frame and therefore not translated.

5 For hybridisation probes, it may be desirable to use nucleic acid analogues, in order to improve the stability and binding affinity. A number of modifications have been described that alter the chemistry of the phosphodiester backbone, sugars or heterocyclic bases.

10 Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulphur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH₂-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire phosphodiester backbone with a peptide linkage.

15 Sugar modifications are also used to enhance stability and affinity. The α -anomer of deoxyribose may be used, where the base is inverted with respect to the natural β -anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity.

20 Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological
25 activity when substituted for deoxythymidine and deoxycytidine, respectively.

The DNA sequences, particularly nucleic acid analogues as described above, may be used as antisense sequences.

30 In accordance with another embodiment of the present invention, there are provided cells containing the above-described nucleic acids. Such host cells such as prokaryote,

yeast and higher eukaryote cells may be used for replicating DNA and producing 5'OT-EST. Suitable prokaryotes include eubacteria, such as Gram-negative or Gram-positive organisms, such as *E. coli*, e.g. *E. coli* K-12 strains, DH5 α and HB101, or Bacilli. Further hosts suitable for 5'OT-EST encoding vectors include eukaryotic microbes such as filamentous fungi or yeast, e.g. *Saccharomyces cerevisiae*. Higher eukaryotic cells include insect and vertebrate cells, particularly mammalian cells, including human cells, or nucleated cells from other multicellular organisms. The propagation of vertebrate cells in culture (tissue culture) is a routine procedure. Examples of useful mammalian host cell lines are epithelial or fibroblastic cell lines such as Chinese hamster ovary (CHO) cells, NIH 3T3 cells, HeLa cells or 293T cells. The host cells referred to in this disclosure comprise cells in *in vitro* culture as well as cells that are within a host animal.

DNA may be stably incorporated into cells or may be transiently expressed using methods known in the art. Stably transfected mammalian cells may be prepared by transfecting cells with an expression vector having a selectable marker gene, and growing the transfected cells under conditions selective for cells expressing the marker gene. To prepare transient transfectants, mammalian cells are transfected with a reporter gene to monitor transfection efficiency.

To produce such stably or transiently transfected cells, the cells should be transfected with a sufficient amount of 5'OT-EST-encoding nucleic acid to form 5'OT-EST. The precise amounts of DNA encoding 5'OT-EST may be empirically determined and optimised for a particular cell and assay.

Host cells are transfected or, preferably, transformed with the above-captioned expression or cloning vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences. Heterologous DNA may be introduced into host cells by any method known in the art, such as transfection with a vector encoding a heterologous DNA by the calcium phosphate coprecipitation

technique or by electroporation. Numerous methods of transfection are known to the skilled worker in the field. Successful transfection is generally recognised when any indication of the operation of this vector occurs in the host cell. Transformation is achieved using standard techniques appropriate to the particular host cells used.

5

Incorporation of cloned DNA into a suitable expression vector, transfection of eukaryotic cells with a plasmid vector or a combination of plasmid vectors, each encoding one or more distinct genes or with linear DNA, and selection of transfected cells are well known in the art (see, e.g. Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press).

10

Transfected or transformed cells are cultured using media and culturing methods known in the art, preferably under conditions, whereby 5'OT-EST encoded by the DNA is expressed. The composition of suitable media is known to those in the art, so that they can be readily prepared. Suitable culturing media are also commercially available.

15

The cDNA or genomic DNA encoding native or mutant 5'OT-EST can be incorporated into vectors for further manipulation. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous DNA into cells for either expression or replication thereof. Selection and use of such vehicles are well within the skill of the artisan. Many vectors are available, and selection of appropriate vector will depend on the intended use of the vector, i.e. whether it is to be used for DNA amplification or for DNA expression, the size of the DNA to be inserted into the vector, and the host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the host cell for which it is compatible. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and optionally a signal sequence.

20

25

30

Both expression and cloning vectors generally contain nucleic acid sequence that enable the vector to replicate in one or more selected host cells. Typically in cloning vectors, this sequence is one that enables the vector to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of bacteria, yeast and viruses. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria, the 2 μ plasmid origin is suitable for yeast, and various viral origins (e.g. SV 40, polyoma, adenovirus) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors unless these are used in mammalian cells competent for high level DNA replication, such as COS cells.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in *E. coli* and then the same vector is transfected into yeast or mammalian cells even though it is not capable of replicating independently of the host cell chromosome. DNA may also be replicated by insertion into the host genome. However, the recovery of genomic DNA encoding 5'OT-EST is more complex than that of exogenously replicated vector because restriction enzyme digestion is required to excise 5'OT-EST DNA. DNA can be amplified by PCR and be directly transfected into the host cells without any replication component.

Advantageously, an expression and cloning vector may contain a selection gene also referred to as selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

As to a selective gene marker appropriate for yeast, any marker gene can be used which facilitates the selection for transformants due to the phenotypic expression of the marker gene. Suitable markers for yeast are, for example, those conferring resistance to antibiotics G418, hygromycin or bleomycin, or provide for prototrophy in an auxotrophic yeast mutant, for example the URA3, LEU2, LYS2, TRP1, or HIS3 gene.

Since the replication of vectors is conveniently done in *E. coli*, an *E. coli* genetic marker and an *E. coli* origin of replication are advantageously included. These can be obtained from *E. coli* plasmids, such as pBR322, Bluescript® vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both *E. coli* replication origin and *E. coli* genetic marker conferring resistance to antibiotics, such as ampicillin.

Suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up 5'OT-EST nucleic acid, such as dihydrofolate reductase (DHFR, methotrexate resistance), thymidine kinase, or genes conferring resistance to G418 or hygromycin. The mammalian cell transformants are placed under selection pressure which only those transformants which have taken up and are expressing the marker are uniquely adapted to survive. In the case of a DHFR or glutamine synthase (GS) marker, selection pressure can be imposed by culturing the transformants under conditions in which the pressure is progressively increased, thereby leading to amplification (at its chromosomal integration site) of both the selection gene and the linked DNA that encodes 5'OT-EST. Amplification is the process by which genes in greater demand for the production of a protein critical for growth, together with closely associated genes which may encode a desired protein, are reiterated in tandem within the chromosomes of recombinant cells. Increased quantities of desired protein are usually synthesised from thus amplified DNA.

Expression and cloning vectors usually contain a promoter that is recognised by the host organism and is operably linked to 5'OT-EST nucleic acid. Such a promoter may be inducible or constitutive. The promoters are operably linked to DNA encoding 5'OT-EST by removing the promoter from the source DNA by restriction enzyme

digestion and inserting the isolated promoter sequence into the vector. Both the native 5'OT-EST promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of 5'OT-EST DNA. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

Promoters suitable for use with prokaryotic hosts include, for example, the β -lactamase and lactose promoter systems, alkaline phosphatase, the tryptophan (*trp*) promoter system and hybrid promoters such as the *tac* promoter. Their nucleotide sequences have been published, thereby enabling the skilled worker operably to ligate them to DNA encoding 5'OT-EST, using linkers or adaptors to supply any required restriction sites. Promoters for use in bacterial systems will also generally contain a Shine-Delgarno sequence operably linked to the DNA encoding 5'OT-EST.

Preferred expression vectors are bacterial expression vectors which comprise a promoter of a bacteriophage such as phage λ or T7 which is capable of functioning in the bacteria. In one of the most widely used expression systems, the nucleic acid encoding the fusion protein may be transcribed from the vector by T7 RNA polymerase (Studier *et al.*, Methods in Enzymol. 185; 60-89, 1990). In the *E. coli* BL21(DE3) host strain, used in conjunction with pET vectors, the T7 RNA polymerase is produced from the λ -lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG inducible *lac* UV5 promoter. This system has been employed successfully for over-production of many proteins. Alternatively the polymerase gene may be introduced on a lambda phage by infection with an *int*- phage such as the CE6 phage which is commercially available (Novagen, Madison, USA). other vectors include vectors containing the lambda PL promoter such as PLEX (Invitrogen, NL) , vectors containing the *trc* promoters such as pTrcHisXpressTm (Invitrogen) or pTrc99 (Pharmacia Biotech, SE) , or vectors containing the *tac* promoter such as pKK223-3 (Pharmacia Biotech) or PMAL (new England Biolabs, MA, USA).

Moreover, the *5'OT-EST* gene according to the invention preferably includes a secretion sequence in order to facilitate secretion of the polypeptide from bacterial hosts, such that it will be produced as a soluble native peptide rather than in an inclusion body. The peptide may be recovered from the bacterial periplasmic space, or the culture medium, as appropriate.

Suitable promoting sequences for use with yeast hosts may be regulated or constitutive and are preferably derived from a highly expressed yeast gene, especially a *Saccharomyces cerevisiae* gene. Thus, the promoter of the TRP1 gene, the ADHI or ADHII gene, the acid phosphatase (PH05) gene, a promoter of the yeast mating pheromone genes coding for the α - or α -factor or a promoter derived from a gene encoding a glycolytic enzyme such as the promoter of the enolase, glyceraldehyde-3-phosphate dehydrogenase (GAP), 3-phospho glycerate kinase (PGK), hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triose phosphate isomerase, phosphoglucose isomerase or glucokinase genes, the *S. cerevisiae* GAL 4 gene, the *S. pombe* nmt 1 gene or a promoter from the TATA binding protein (TBP) gene can be used. Furthermore, it is possible to use hybrid promoters comprising upstream activation sequences (UAS) of one yeast gene and downstream promoter elements including a functional TATA box of another yeast gene, for example a hybrid promoter including the UAS(s) of the yeast PH05 gene and downstream promoter elements including a functional TATA box of the yeast GAP gene (PH05-GAP hybrid promoter). A suitable constitutive PH05 promoter is e.g. a shortened acid phosphatase PH05 promoter devoid of the upstream regulatory elements (UAS) such as the PH05 (-173) promoter element starting at nucleotide -173 and ending at nucleotide -9 of the PH05 gene.

5'OT-EST gene transcription from vectors in mammalian hosts may be controlled by promoters derived from the genomes of viruses such as polyoma virus, adenovirus, fowlpox virus, bovine papilloma virus, avian sarcoma virus, cytomegalovirus (CMV), a retrovirus and Simian Virus 40 (SV40), from heterologous mammalian promoters such

as the actin promoter or a very strong promoter, e.g. a ribosomal protein promoter, and from the promoter normally associated with *5'OT-EST* sequence, provided such promoters are compatible with the host cell systems.

- 5 Transcription of a DNA encoding *5'OT-EST* by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are relatively orientation and position independent. Many enhancer sequences are known from mammalian genes (e.g. elastase and globin). However, typically one will employ an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270) and the CMV early promoter enhancer. The enhancer
10 may be spliced into the vector at a position 5' or 3' to *5'OT-EST* DNA, but is preferably located at a site 5' from the promoter.

- Advantageously, a eukaryotic expression vector encoding *5'OT-EST* may comprise a locus control region (LCR). LCRs are capable of directing high-level integration site independent expression of transgenes integrated into host cell chromatin, which is of importance especially where the *5'OT-EST* gene is to be expressed in the context of a permanently-transfected eukaryotic cell line in which chromosomal integration of the vector has occurred, in vectors designed for gene therapy applications or in transgenic
15 animals.
20

- Eukaryotic expression vectors will also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These
25 regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding *5'OT-EST*.

- An expression vector includes any vector capable of expressing *5'OT-EST* nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that
30 are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or

other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding 5'OT-EST may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, *et al.*, (1989) NAR 17, 6418).

Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding 5'OT-EST in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of 5'OT-EST. For the purposes of the present invention, transient expression systems are useful e.g. for identifying 5'OT-EST mutants, to identify potential phosphorylation sites, or to characterise functional domains of the protein.

Construction of vectors according to the invention employs conventional ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing 5'OT-EST expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

In a further aspect, the present invention provides a transgenic non-human animal which

expresses, as a result of transformation with a transgene, 5'OT-EST or a mutant thereof as defined herein. Preferred animals include mammals, especially rats.

5 Preferably, the non-human animal is a mammal suitable for use as a test system for therapies and treatments relating to obesity, including human obesity and animal obesity, which is of concern in household animals such as cats and dogs. Thus, the mammal may be a cat or a dog, or other household pet; it is preferably a rodent, such as a mouse or a rat, particularly a rodent adapted for laboratory testing whose genotype and general characteristics are well known.

10 Any technique may be used to generate transgenic animals according to the invention. Preferably, the technique involves transfer of a transgene comprising 5'OT-EST to the pronucleus of a single-cell embryo, prior to implantation of the embryo into a pseudopregnant foster mother. Such techniques have the advantage that germ-line
15 transgenic animals are readily produced.

Alternatively, transgenic animals may be created by ES cell transfer techniques. In such techniques, ES cells are transformed with the desired transgene and then used to reconstitute an embryo. Animals created by such techniques are normally chimeric for the
20 transgene. However, more accurate positional insertion of the transgene is possible, and selective deletion of endogenous genes by homologous recombination is facilitated (Mansour *et al.*, 1989).

Further techniques include targeted or non-targeted delivery of genes to whole animals,
25 using viral or non-viral vectors. For example, genes may be delivered by recombinant retroviruses or adenovirus vectors, including adeno-assisted virus vectors, which are capable of integrating into the genome of the animal and expressing the delivered gene. Non-viral vectors include liposomal vectors, antibody-targeted DNA-protein complexes and the like.

30 As used herein, "transgenic" animals include animals from which 5'OT-EST has been deleted, as well as animals to which a 5'OT-EST transgene has been added. Optionally,

the endogenous *5'OT-EST* may be deleted, and a transgene bearing a heterologous or homologous *5'OT-EST* gene, which may be wild-type or mutated, inserted into the animal.

- 5 Preferred vectors for creating transgenic animals include linearised naked DNA from a variety of sources. In a preferred embodiment, transgenes may be derived from linearised cosmid sequences, from which the phage-related sequences have optionally been removed.
- 10 The *5'OT-EST* sequences used in a transgene according to the present invention may be inserted separately, or together with further sequences, including reporter genes, further effector genes and the like. Preferably, *5'OT-EST* is comprised in a nucleic acid fragment which comprises the natural wild-type environment of *5'OT-EST*, including flanking sequences.
- 15 *5'OT-EST* is located proximal to the vasopressin (AVP) and oxytocin (OT) genes in the genome, being transcribed in opposite directions from positions closely linked in a single locus. *5'OT-EST* lies 5' of the OT gene in at the OT/AVP locus. Accordingly, the transgene preferably consists of the OT/AVP locus, including *5'OT-EST*.
- 20 Advantageously, one or more of the OT, AVP and *5'OT-EST* genes may be mutated, for example by insertion of reporter genes, such as the hGH gene.

In a highly preferred aspect, the transgene is cosmid cVO14 as described in Figure 4 herein. The complete sequence of cVO14 is set forth in SEQ. ID. No. 17.

25 Transgenic animals according to the invention may comprise single copies of the transgene, or may comprise multiple integrated copies, which may be present as concatamers. Preferably, transgenic animals according to the invention comprise four or more copies of the transgene.

30 Transgenic animals according to the invention may be employed for a variety of purposes. The characteristics of male specificity, central distribution of adiposity, late onset and

severity, and associated morbidity have parallels in the description of several human forms of central obesity. These include, but are not limited to, the condition known as metabolic syndrome, or Syndrome X, as well as other forms of central obesity which may be most severely expressed in human males, with or without reduced fertility and which are associated with increased morbidity. (For recent reviews of the importance of clinical and health care issues in obesity see Science 1998, vol. 280 pp. 1364-1390). Transgenic animals expressing 5'OT-EST or mutants thereof thus have particular beneficial utility as a novel animal model of late-onset human visceral obesity, preponderant in males.

Moreover, the induction of the SLOB phenotype in juvenile rats as a result of dietary fat increases suggests that transgenic animals expressing 5'OT-EST are a model for juvenile obesity in mammals, predominantly male mammals, which is induced by the consumption of a high-fat diet.

Furthermore, the onset of obesity in ovariectomised SLOB female rats suggests the model may be suitable to investigate post-menopausal obesity in female mammals.

For instance, one recognised value of animals bearing 5'OT-EST or mutant constructs is to use such animals and their nontransgenic littermates as animal experimental models for studying obesity or male infertility and their related conditions. Using the information disclosed herein, it is possible to identify transgenic animals before they become obese or sexually mature, and to use them as a model for studying the factors that affect the development of obesity or male infertility in any animal classified as a mammal, including humans, domestic, and farm animals, and zoo, sports, or pet animals, such as but not limited to sheep pigs, cows, horses, dogs, cats, etc.

In particular, rodent models of obesity or infertility are of value in testing the ability of pharmaceutical preparations of novel agents, to be beneficial in delaying or preventing the occurrence, development, course, severity, progression, or exacerbation of obesity or infertility (Mathe, 1995; Fan *et al.*, 1997). Animals bearing 5'OT-EST or mutant constructs are particularly useful in testing agents in this regard, since the phenotype is predictable and non-transgenic littermates are ideal controls.

In addition to screening for unknown compounds, animals bearing 5'OT-EST or mutants thereof may be particularly useful in studies employing administration of natural or recombinant proteins, peptides or other agents or their derivatives already known or suspected to be involved in some forms of obesity or male infertility (e.g. growth hormones, or reproductive hormones, their homologues, analogues, antagonists, inhibitors or secretagogues, or leptin, its homologues, analogues and antagonists) or other natural or pharmacological agents already known to be active and/or of therapeutic value in these conditions (e.g. insulin, thiazolidinones, catecholamines, gonadal steroids) or agents already known to affect their actions, distribution, catabolism or elimination).

Typically in such studies, compounds may be administered to animals bearing 5'OT-EST or mutants thereof and their non-transgenic littermates by oral, parenteral (e.g., intramuscular, intraperitoneal, intravenous, or subcutaneous injection or infusion, or implant), nasal, pulmonary, rectal, sublingual, or topical routes of administration, and can be formulated in dosage forms appropriate for each route of administration, e.g. in soluble form, suspension, or other suitable pharmaceutical formulations.

For example, the effects of such compounds on the obese phenotype may be assessed by carcass analysis, measurement of growth, body weight, body fat distribution, as well as other measures of analytes in body fluids or tissues relevant to obesity (Mathe, 1995; Shillabeer, 1992). These include, but are not limited to, cholesterol, triglycerides, fatty acids, lipoproteins, and other dietary constituents or metabolites, as well as metabolic hormones, such as leptin, insulin, glucagon, catecholamines or glucocorticoids. Other relevant parameters include cardiovascular measures (Reaven, 1988, Gray & Yudkin, 1997). These may include measures of systolic or diastolic blood pressure, cardiac output, or vascular resistance, together with morphological changes to organ systems known to be affected by cardiovascular or obesity disorders, such as heart, major or minor blood vessels, their muscle or endothelial layers, and their elasticity or fragility. See for example McNamee *et al.* (1994).

Similarly, parameters related to the infertility phenotype that may be measured, include,

but are not limited to, testicular weight, volume, development, spermatogenesis, sperm number, motility or ability to fertilise oocytes. They may also include measures of testicular fluid production and constituents, as well as products of other accessory organs including seminal vesicles or prostate, as well as hormones, receptors, and proteins important in male sexual function, such as testosterone, LH, FSH, inhibin or activin. Other responses that may be affected include energy expenditure, physical activity, ingestive behaviour, excretory behaviour, or reproductive behaviour, or the organs, hormones or receptors commonly recognised to be associated with these physiological systems, their metabolism or morphological structure.

10

Compounds identified as effective in such screening or analysis based on the use of animals bearing 5'*OT-EST* or mutants thereof are particularly useful in treatment of late-onset visceral obesity, or male infertility, in particular where they occur in combination, and disorders related to these conditions with a view to delaying or preventing the occurrence, development, course, severity or progression of the phenotype, avoiding its exacerbation, and preferably promoting its amelioration or cure in animals of commercial importance, or more preferably in humans.

15

In another embodiment, converse but also therapeutically valuable compounds may be developed based on screening or analysis as above in animals bearing 5'*OT-EST* or mutants thereof but which are intended to *promote* the occurrence, development, or progression of *increased* fat deposition or increased calorie intake or decreased energy consumption. Such disorders in humans include, but are not limited to, wasting, or anorexia, or cachexia, associated with prolonged illness, or malabsorptive states or catabolic states associated with other diseases, such as, but not limited to, inflammatory conditions, Crohns disease, or AIDS wasting, or burns, or cancer, or bone disease.

20

25

Similarly, therapeutically valuable approaches may be developed based on screening or analysis as above in animals bearing 5'*OT-EST* or mutants thereof but which *reduce* the degree of male fertility in those conditions in which it might be beneficial. For example, this may be beneficial in the control of populations in animals of commercial or environmental importance, or to develop novel forms of contraception which may be

30

effective in human males. Such approaches specifically include, but are not limited to, the possibility of blocking 5'OT-EST or mutants thereof function by administering 5'OT-EST mutant products or by immunisation against 5'OT-EST to generate neutralising antibodies that interfere with the normal functioning of this gene product in the testis, or hypothalamus or adrenal gland or gastrointestinal tract, or other organ system in which 5'OT-EST is expressed or upon which its products act.

The development and late-onset of obesity in transgenic animals may be particularly useful in studying the chronic effects of novel food additives or formulations designed to prevent or exacerbate the deposition of fat in animals of commercial importance, destined for use in human food products or dietary aids. Such compounds may be administered as above and their effects on the development, course, severity, progression, exacerbation, amelioration or cure of the obesity phenotype assessed as described above. Additives or formulations shown to reduce the development of visceral obesity in this model may have utility in human food products or dietary aids or find beneficial medicinal use in reducing fat accretion or retention.

In another embodiment, transgenic animals, such as mice bearing 5'OT-EST or mutants thereof or in which 5'OT-EST has been disrupted, may be usefully intercrossed with other animal strains with defined mutations, or with undefined genetic backgrounds associated with propensity for the development of obesity or infertility. Comparison of the resulting progeny with or without the 5'OT-EST transgene will provide additional information on the alterations in occurrence, development, course, severity, progression, exacerbation, amelioration or cure of the obesity phenotype when expressed in these other genetic backgrounds, and analysed as described above. Such intercrossing may then be envisaged to enhance the utility of the resulting progeny exhibiting the obesity phenotype.

Examples of this use include (without being limited to) interbreeding with Zucker fa/fa rats (Iida *et al.*, 1996), corpulent (cp) rats (Kahle *et al.*, 1997), OLETF rats (Takiguchi *et al.*, 1998), ZDF rats, tfm rats, spontaneously hypertensive or salt-sensitive rats (Michaelis *et al.*, 1995) or other dwarf rats such as dw/dw (Charlton *et al.*, 1988) or dr/dr rats (Takeuchi *et al.*, 1991). An example of the utility of this approach is given by

(Michaelis *et al.*, 1995). Examples of mouse lines that may be usefully interbred with mice carrying transgenes or deletions affecting *5'OT-EST* include, but are not limited to, ob/ob, db/db, tfm/tfm or hpg/hpg mice. A related example includes intercrossing mice carrying transgenes or deletions affecting *5'OT-EST* with other strains of mice in which genes already known to be involved in obesity or male fertility have been deleted by homologous recombination or introduced by transgenesis (singly or in combination). Examples of these are already known to include (but are not limited to) leptin, tubby and related genes, NPY, insulin, GLP-1, IGF-1, IGF-II, MCH, CRH, POMC, CCK, orexins or hypocretins, CART peptides, agouti protein, as well as the genes or alternate products structurally related to or homologous with, the above peptides. This example is also intended to include mice with disruptions in or extra copies of normal or mutated forms of, genes for the specific receptors of the peptides listed above (for example NPY receptors, such as subtype 5), or bombesin-receptor 3, IRS-1 or 2, uncoupling proteins such as UCP1-3, carboxypeptidase E, or PPARs or adrenergic receptor subtype 3 or TNF alpha or, all of which have been implicated in obesity.

Similarly, the fertility disruption in transgenic animals according to the invention may also be studied to advantage by crossing these animals or other animals in which *5'OT-EST* has been disrupted onto genetic backgrounds in which genes for gonadal or adrenal steroid biosynthesis or metabolism or gonadal steroid receptors and other reproductive hormones or their receptors or hypothalamic or pituitary hormones thought to affect male fertility (such as gonadotrophins, activins, inhibins, PRL, GnRH or transcription factors such as DAX1 or SF1, or other known gene products affecting male gonadal development, such as MIS, AMH, SCF) have been disrupted. Comparison of the resulting progeny with or without the *5'OT-EST* or mutant transgene may shed light on the alterations in occurrence, development, course, severity, progression, exacerbation, amelioration or cure of the infertility phenotype when expressed in these other genetic backgrounds. Such intercrossed lines, for example with those genetic strains as outlined above, and in which the obesity phenotype is present in full or in a modified form, may also be additionally useful for screening applications.

Conversely, the transfer of the obese phenotype onto these genetic backgrounds may also

alter the occurrence, development, course, severity, progression, exacerbation, amelioration or cure of the specific phenotypes expressed in the strain with which animals bearing *5'OT-EST* or mutants thereof are bred. Such intercrossed lines, for example with those genetic strains as outlined above, and in which the obesity phenotype is present in full or in a modified form, may also be useful for screening applications.

Animals bearing *5'OT-EST* or mutants thereof may be used to study the transfer of other gene products other than by breeding, e.g. by administration of suitable vectors containing constructs expressing proteins of interest, or by transgenesis. Such examples include, but are not limited to, constructs containing the gene products or analogues of other genes already thought to be active in obesity or male infertility, whose effects may be advantageously studied in transgenic animals according to the invention due to the predictable development of their phenotype. Such genes and their products include those mentioned in above in relation to alternative obese strains. Such derived animals in which the obesity phenotype is present in full or in a modified form, may also be useful for the various applications outlined above.

Animals bearing *5'OT-EST* or mutants thereof and exhibiting a specific late-onset visceral obesity may prove of particular value when used in a similar way to screen for the beneficial effects of reducing or eliminating other gene products by their silencing or elimination as described above using transgenesis, or homologous recombination, or by adenoviral delivery of antisense nucleotides. Examination of any alterations in the occurrence, development, course, severity, or progression of the obesity phenotype in these genetic backgrounds would be of utility in identifying the role, if any, of such disrupted genes in the expression of the obesity phenotype. Such animals in which the obesity phenotype is present in full or in a modified form, may also be useful for the applications outlined above.

In another embodiment, animals bearing *5'OT-EST* or mutants thereof or material derived from them and/or their nontransgenic littermates may prove useful in experiments designed to identify obesity -related or male-fertility-related differences in gene expression, RNA transcripts, proteins, or other biochemical measures, such as, but not

limited to lipids, peptides, carbohydrates, amino acids or compounds or precursors or metabolites thereof, or their distribution, in whole animals, or in samples of biological fluids taken from animals bearing *5'OT-EST* or mutants thereof. These may include, but are not limited to: serum, plasma, lymph fluid, synovial fluid, follicular fluid, seminal
5 fluid, amniotic fluid, milk, whole blood, urine, cerebrospinal fluid, saliva, sputum, tears, perspiration, mucus, tissue culture medium, tissue extracts, and cellular extracts.

Similar analyses may be advantageously performed in samples of any tissue from animals bearing *5'OT-EST* or mutants thereof or their non-transgenic littermates, or tissue derived
10 from animals interbred with SLOB rats or other animals bearing *5'OT-EST* or mutants thereof. Such tissues are preferably (but not limited to) endocrine tissues, such as pancreas, adrenals, or pituitary gland, adipose tissues from different locations, preferably but not limited to, inguinal, omental, perirenal, subcutaneous, mammary, periorbital or other regions, thermogenic fat, brown or white adipose tissue in other locations, areas of
15 the CNS though to be involved in obesity, preferably but not limited to the hypothalamus, and other tissues, preferably, but not restricted to liver, gastrointestinal tract, gonads, heart, musculoskeletal system, immune system, kidney, connective tissue including skin, epithelial or endothelial tissues.

20 Specifically included are cells or tissues removed from animals bearing *5'OT-EST* or mutants thereof, or animals interbred with them, and maintained thereafter ex vivo, e.g. in tissue culture, or by transplantation in animals bearing *5'OT-EST* or mutants thereof or other hosts, with or without immune suppression, provided the particular utility is enhanced by the presence of the obesity gene or phenotype.

25

The invention thus provides the use of a tissue derived from a transgenic animal according to any one of claims 17 to 24 in a screen to identify a genetic cause of obesity, comprising the steps of:

- a) isolating one or more gene products from tissue derived from a transgenic animal
30 according to any one of claims 17 to 24; and
- b) determining whether the expression of a gene product is correlated with obesity.

Tissues derived from SLOB rats, including in particular fat pad tissue, may for example be used for differential screening in order to determine differences in gene expression between obese and non-obese animals. The gene products analysed may be nucleic acid or protein gene products. For example, mRNA may be isolated from the tissue and
5 screened to identify differentially expressed transcripts. In particular, gene products which may be involved in cellular lipid transport may be identified by such means.

The development of obesity or male infertility itself in animals bearing *5'OT-EST* or mutants thereof is predicted to induce secondary changes in other obesity or fertility -
10 related parameters and regulators. These include, but are not limited to, blood pressure, pituitary hormones, sperm development, maturation, and/or motility, lipid mobilising enzymes or receptors, or agents controlling these. The latter include, but are not limited to leptin and its receptors, melanocortin, NPY, catecholamines, adrenal or gonadal or
15 pituitary hormones, gut hormones such as insulin and glucagon, growth hormone and other growth factors such as members of the GH and IGF-1 families, their binding proteins and receptors. They may also include drugs of several classes that have been thought useful in obesity. Examples of such classes include agents affecting the serotonin system or the fat cell free fatty acid uptake or release or metabolism or lipase activity or
20 hepatic lipid uptake, or insulin sensitisers. This example may also include morphological alterations in any tissue or cells of the cardiovascular system, including but not limited to, the heart and major blood vessels, other blood vessels carrying either arterial or venous blood, and any or all cells comprising these tissues.

Transgenic animals according to the invention appear to present with obesity without
25 obvious diabetes or hypercortisolism. They may thus prove particularly beneficial in studying the developmental changes in these secondary parameters induced by other means, in the development of diabetes, or hypertension or cardiovascular disease or hypercortisolism (Russell *et al.*, 1993), all which are known to be associated with obesity in humans (Reaven, 1988). Examples of such means includes (but is not limited to)
30 variation in dietary components or quantity, or treatment with diabetogenic agents, such as GH or cortisol. Examples of such agents affecting cardiac output or peripheral resistance or blood pressure include angiotensin-converting enzyme inhibitors or cardiac

glycosides, or beta adrenergic receptor 3 agonists or antagonists.

Morphological changes may also be seen in adipose tissues or cells, or the other tissues in the body containing fat, such as the liver and related cells, or the skeleton, or in other organs or tissues in the gonadal system, relating to the effects on male fertility. Differences in these measures detected specifically in animals bearing *5'OT-EST* or mutants thereof and their alteration by elimination, blockade, endogenous stimulation, or exogenous administration of anti-obesity or other agents affective in obesity or male infertility or related disorders would provide novel approaches to evaluate improve and perfect existing or novel therapeutic approaches to obesity or male infertility in other animals of commercial importance, and more preferably, in humans. An obvious example is the ready source of adipocyte cells and products from specific fat depots that are differentially increased in transgenic animals according to the invention. Responses to agents affecting fat cell metabolism or fat storage or lipogenesis or lipolysis or lipid-lowering agents, may be studied with particular advantage to discern effects on visceral or peripheral fat tissues. and to seek differential effects on fat from different depots in the animals.

The information disclosed herein will enable those skilled in the art to produce protein or peptide fragments corresponding in sequence to *5'OT-EST* or mutants thereof, as described above. Such proteins or peptides (or simple analogues thereof), when administered to rats or other animals, or more preferably humans, would be expected to affect the development of obesity and fertility in males, and serve as the basis for the development of similar compounds, based on the homologous human sequences, useful in the treatment of these conditions in humans.

This also includes simple analogues that incorporate alterations known to improve the in vivo stability or delay the clearance, elimination or metabolism of proteins or peptides such as those derived from *5'OT-EST* or mutants thereof. Such alterations are obvious to those skilled in the art, and examples include amidation or acetylation of C or N-termini of peptides, or replacement of methionine residues with norleucine residues to avoid oxidation. This example also includes formulations or modifications of proteins known

to be effect for the same purposes, e.g. by PEGylation to prolong the half-life of peptides or proteins, or formulations of proteins with inert carriers (such as mannitol or lactose) or buffers or salts, that provide stable solutions suitable for in vivo administration of the active agents to animals or to humans.

5

In another embodiment, the information disclosed herein will enable those skilled in the art to design nucleotide probes for, or develop polyclonal or monoclonal antibodies against, the DNA, RNA or protein sequences corresponding to the whole or parts of the 5'OT-EST gene in other animals of commercial importance, or more preferably, humans.

10 These are of value in diagnostic tests to screen for mutations in this gene in animal or human populations subject to variations in obesity or fertility. They may also be used to monitor the development, progression, amelioration or cure of obesity or infertility as may be reflected in changes in the activity of this gene or its products. Such predictive tests are recognised to have beneficial value when applied to the human population (Whitaker *et*
15 *al.*, 1997)

Examples of such probes or peptides or proteins used to develop antibodies include those predicted from the wild-type and mutated sequences in the rat 5'OT-EST gene and mutants thereof, as well as their derivatives as described above, as well as those that may
20 be inferred from homologous genes in human and mouse, either as intact sequences or formed in whole or in part as fusion sequences with other proteins to facilitate production or purification by standard methods known to those skilled in the art. For this purpose, products of the 5'OT-EST gene may also be reacted with, produced as fusion products with, or mixed with, other proteins or adjuvants known to enhance the immune response.

25 Also included are modifications to the nucleotide sequences, already known to those skilled in the arts to confer useful chemical properties on the products, for example by incorporating modified nucleotides to render them more stable, or to incorporate nucleotides tagged with functional groups such as biotin or digoxigenin, or incorporation of fluorescent or radioisotopically labelled derivatives, in order to render the products
30 themselves more readily detectable.

In a further embodiment, the information disclosed herein will enable those skilled in the

art to isolate factors which interact directly with 5'OT EST or mutants thereof. For example, two-hybrid screens provide a means for isolating genes for proteins which interact with 5'OT EST or mutant proteins, their fragments or derivatives. Similarly, coprecipitation studies using antibodies directed against 5'OT EST or mutants thereof, produced as outlined herein, might allow identification of such interacting factors. Such factors, when administered to rats or other animals, or more preferably humans, would be expected to affect the development of obesity and/or male infertility in a similar fashion to that seen in transgenic animals, and would therefore be predicted to have similar uses. The use of transgenic animals or materials derived from them, or from the information disclosed herein, has particular utility in providing a specific means of isolating such factors that interact directly with the novel gene products disclosed herein, as well as in screening their biological activities in vivo.

The invention is described further, for the purpose of illustration only, in the following examples.

MATERIALS AND METHODS

Bacterial cultures.

All media are made with double distilled water and autoclaved prior to use. Liquid cultures of bacteria are incubated with shaking at 37 °C in either LB broth or terrific broth. Bacterial colonies are grown on agar plates made with either LB broth or terrific broth with 15g/l bacto-agar. These media are supplemented with combinations of 20µg/ml or 50µg/ml ampicillin, 20µg/ml tetracycline and 0.2% glucose. Bacterial clones are stored at - 80 °C after the addition of 15% glycerol.

Purification of nucleic acids.

Aqueous solutions containing DNA are purified by vortex mixing with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). The emulsion is then centrifuged at 12,000 rpm for 5 minutes in a microfuge at room temperature. DNA is precipitated by adding 3M sodium acetate (pH 5.2) to a final concentration of 300mM and two volumes

of absolute ethanol. The samples are frozen before centrifugation, the supernatant removed and the pellet resuspended in 10mM Tris.HCl pH 8, 1mM EDTA (TE buffer).

DNA preparation from bacteria stocks.

5 The alkaline lysis method of DNA isolation may be used (Birnboim and Doly, 1979; Sambrook *et al.*, 1989) to prepare plasmid DNA from small volumes of bacterial cultures, typically 10ml. For large scale preparation of plasmid and cosmid DNA, DNA may also be prepared from 1 L overnight cultures by the alkaline lysis method. DNA is dissolved in 100mM Tris.HCl pH 8, 1mM EDTA and further purified on a caesium
10 chloride gradient which is centrifuged at 55,000rpm overnight (Sambrook *et al.*, 1989).

Preparation of genomic DNA from animal tissue.

 Rat tail biopsies, up to 1 cm, are taken from 10-14 day old rats and placed in 50mM Tris.HCl pH8, 100mM EDTA, 100mM NaCl ('tail mix'). Genomic DNA is
15 prepared following a standard procedure (Hogan *et al.* 1986) involving incubation with proteinase K, RNase A, phenol extraction and precipitation with isopropanol. Genomic DNA from other tissue such as liver may be prepared by the same method, though this requires additional homogenisation in a larger volume (typically 5ml) of tail mix using a Kinematica Polytron PT 3000 homogeniser prior to the preparation of DNA from a
20 smaller aliquot of homogenate.

Restriction digestion of DNA.

 Restriction enzyme digestion is performed using standard procedures in accordance with manufacturers instructions. Enzymes are sourced from Boehringer
25 Mannheim, Cambio, or New England Biolabs. Plasmid DNA is incubated for up to 4 hours whilst genomic DNA digests may be incubated overnight.

Subcloning DNA fragments into plasmid vectors.

Blunting of DNA fragments with a 3' overhang

30 After digestion of DNA with a restriction enzyme which leaves a 3' overhang, the overhang may be removed by incubation with T4 DNA polymerase to create a blunt end for ligation with other blunt-ended DNA fragments. The digests are phenol-extracted,

ethanol-precipitated with the addition of 10 μ g tRNA, and resuspended in TE. MgCl₂ and deoxynucleotide triphosphates (dNTPs) are added to final concentrations of 10mM and 0.1mM respectively prior to the addition of 2 units of T4 DNA polymerase (New England Biolabs). The reaction is incubated for 15 minutes at 12 °C. The polymerase is inactivated at 75 °C for 10 minutes before purification.

Blunting a DNA fragment by refilling the 5' overhang.

DNA fragments with 5' overhangs may be blunted by filling in the single stranded ends. This may be done using the Klenow fragment of E. coli DNA polymerase I (New England Biolabs). The DNA is digested with an appropriate restriction enzyme, phenol extracted, ethanol precipitated with the addition of 1 μ g of tRNA and resuspended in TE. 10x Klenow buffer (0.5M Tris.HCl, pH7.6, 0.1M MgCl₂) and dNTPs to a final concentration of 1x and 0.2mM respectively are added with 10 units of Klenow fragment. The reaction is incubated at 37 °C for 30 minutes prior to purification of the DNA.

Vector dephosphorylation.

Calf alkaline phosphatase (CAP) may be used to remove 5' phosphate groups from digested vectors to prevent self-ligation during subcloning. Plasmid and cosmid vectors, linearised with restriction enzymes, are incubated with 2 units of CAP (Boehringer) in 50mM Tris.HCl, pH 8.5, 50 mM EDTA, for 30 minutes prior to purification.

Inserting linkers into DNA fragments.

Digested plasmid DNA may be blunted if necessary, phenol-extracted, ethanol-precipitated with the addition of 1 μ g of tRNA and resuspended in TE. 0.5-1 μ g of phosphorylated linkers are ligated to linearised, blunt-ended plasmid DNA. Ligations are performed in a final concentration of 1x ligase buffer (50mM Tris.HCl (pH 7.5), 10mM MgCl₂, 10mM dithiothreitol, 1mM ATP, 25 μ g/ml bovine serum albumin (BSA), 0.5mM spermidine-HCl with the addition of 400units of T4 DNA ligase (New England Biolabs). The reactions are incubated at room temperature overnight. The enzyme is then inactivated at 65 °C for 15 minutes. The linkered fragments are digested with an excess amount of the appropriate restriction enzyme and the DNA purified prior to further subcloning procedures.

Electrophoresis of DNA fragments

DNA fragments may be electrophoresed in gels of varying percentages of agarose in 90mM Tris-borate, 2mM EDTA, pH 8.0 (TBE buffer) containing 0.5ng/ml ethidium bromide. The DNA bands are visualised on an ultraviolet transilluminator and photographed. Size markers may be Lambda DNA digested with Bst EII, pUC19 DNA digested with Msp I, or a commercially available 1kb ladder (Gibco-BRL).

Purification of DNA fragments.

Digested, blunted, dephosphorylated or linker DNA fragments are electrophoresed in low melting-point agarose. Gel bands are excised, melted at 65 °C for 5 minutes, and extracted twice with phenol/0.3M NaOAc. Following a phenol extraction and ethanol precipitation with the addition of 1µg tRNA, the DNA is recovered by centrifugation and resuspended in TE. Alternatively, DNA fragments may be purified from agarose using the Prep-A-Gene DNA Purification System (Bio-Rad Laboratories) according to manufacturers instructions.

Purification of larger cosmid-containing fragments.

Large vectors are digested and treated with 50 units of CAP for in excess of 3 hours. EDTA and SDS are added to final concentrations of 5mM and 0.5% respectively. The phosphatase is denatured for 1 hour at 65 °C and the solution is phenol-extracted, ethanol precipitated with the addition of 1µg tRNA, and the DNA recovered is resuspended in TE.

Ligation of DNA fragments into phosphatased vectors.

After purification, DNA fragments and vectors are mixed at equimolar ratios at an approximate concentration of up to 80ng/ml, whilst DNA for recircularisation is used at a concentration of 20ng/ml. Ligation may be done in a volume of 5µl with 200 units of T4 DNA ligase (New England Biolabs) in ligase buffer. Two control reactions are performed simultaneously omitting the insert alone, or omitting both insert and ligase. Ligations are incubated overnight at 16 °C.

Preparation of competent cells.

	psi-broth	5g/l bacto-yeast extract, 20g/l bacto-tryptone, 5g/l MgSO ₄ , adjusted to pH 7.6 with NaOH.
5	TFbI	30mM KAc, 100mM KCl, 10mM CaCl ₂ , 50mM MnCl ₂ , 15% glycerol (v/v), adjusted pH 5.8 with acetic acid and filter sterilised.
10	TFbII	10mM PIPES, 74mM CaCl ₂ , 10mM KCl, 15% glycerol (v/v), adjusted to pH 6.5 with acetic acid and filter sterilised.

Competent cells yielding a transformation frequency $>5 \times 10^8$ transformed colonies per μg of supercoiled plasmid DNA may be prepared by a method modified from Hanahan *et al.* (1983). Bacteria of the strain DH10B (Grant *et al.*, 1990) are plated on an agar plate and grown overnight at 37 °C. 10ml of psi-broth is then inoculated with 4 colonies from this plate. The bacteria are then shaken at 37 °C until OD550=0.3.

5 ml of this broth is then diluted into 100 ml psi-broth and shaken until OD550=0.28. The flask is then placed on ice, the bacteria are centrifuged at 4 °C for 15 minutes at 2,000 rpm. The supernatant is removed and the pellet allowed to dry briefly before being resuspended in 20ml TFbI. This suspension is left on ice for 5 minutes and then centrifuged at 2,000 rpm for 10 minutes at 4 °C. The supernatant is then removed and the pellet resuspended in 3 ml TFbII and placed on ice for 15 minutes. Aliquots are then frozen on dry ice and stored at -80 °C. The competence of the cells may be tested by transforming plasmid DNA of known concentrations.

Transformation of competent cells.

Competent cells are thawed on ice before 50 μl of cells is added to each ligation. These tubes are then incubated on ice for 30 minutes. The cells are then subjected to heatshock at 42 °C for 90 seconds before being placed on ice for 2 minutes. 0.4ml of LB

broth is added and the culture is shaken at 37°C for 1 hour. Cells are then incubated overnight at 37 °C on agar plates containing the appropriate antibiotic. Single colonies are picked with a flamed wire loop and used to inoculate 10ml of media for miniprep of plasmid DNA.

5

Packaging of Cosmid DNA into bacteriophage particles.

Cosmid constructs may be packaged into bacteriophage particles using Gigapack II packaging extracts (Stratagene) and *E. coli* strain DH10B is infected in accordance with manufacturer's instructions.

10

Southern blotting.

DNA (10µg of genomic DNA or 0.5 µg of plasmid DNA) is digested with appropriate restriction enzymes and electrophoresed on agarose gels with marker DNA of known size. After photography, gels are treated as described by Sambrook *et al.* (1989) and the DNA transferred from the gels onto nitro-cellulose filters by the capillary transfer method (Southern, 1975; Sambrook *et al.*, 1989) and these are then baked for 2 hours.

15

Radiolabelling of DNA fragments for Southern blots.

DNA probes are obtained by gel purifying appropriate fragments from restriction digests of subcloned DNA. The DNA is denatured by being incubated for 3 minutes in a boiling water bath. The resulting single-stranded DNA fragments are radiolabelled with α -³²PdCTP, e.g. by the random primer labelling kit, Prime-It II supplied from Stratagene, in accordance with manufacturer's instructions. The labelling reaction is halted by the addition of TES buffer to a final concentration of 10mM Tris.HCl (pH 7.5), 10mM EDTA, 0.1% SDS. Radiolabelled DNA probes are separated from unincorporated nucleotides by eluting through a column containing Sephadex G-50.

20

25

Hybridisation of Southern blots.

30 100X Denhardts solution

2% BSA, 2% Ficoll 400, 2%

Polyvinyl Pyrollidine.

- 1M sodium phosphate buffer (pH6.6) 352 ml 1M Na₂HPO₄, 648ml 1M NaH₂PO₄.
- Prehybridisation mix 0.1mg/ml tRNA, 5x SSC
 50mM Na Phosphate buffer (pH 6.6)
 10x Denhardts solution, 1% SDS.
- Hybridisation mix Prehybridisation mix with the above
 described radiolabelled DNA probe.
- 10 Filters from Southern blotting are gently shaken at 65 °C in prehybridisation mix for a minimum of 2 hours. This solution is then replaced with hybridisation mix and incubated overnight. The filters are washed in varying concentrations of SSC with 0.1% SDS for varying amounts of time dependent on the DNA probe being used. Filters are then dried and placed between two intensifying screens at -70 °C with Kodak "Xomat-AR" film.
- 15 **Screening a rat cosmid library**
- Duplicate filters from a Wistar rat cosmid library containing genomic DNA inserts in the pWE15 cosmid vector (Wahl *et al.* 1987) are prehybridised and hybridised with probes for the rat OT and AVP genes. Following overnight hybridisation, filters are washed with
- 20 3x SSC/0.1% SDS for 20 minutes and then SSC/0.1% SDS twice for 20 minutes. Filters are briefly washed in 2x SSC, dried and autoradiographed. Duplicate hybridisation signals are aligned with the master filters and bacteria are picked, placed in media and left to diffuse. The resulting cultures are grown on terrific broth agar with 20µg/ml ampicillin and replica plated. Replica plating of the bacterial culture from the library screening may
- 25 be performed as previously described (Sambrook *et al.*, 1989). Replica filters are prehybridised and hybridised as above. Positively-hybridising colonies are picked from the master filters and grown in larger volumes of ampicillin-supplemented media for miniprep and southern blot analysis of the cosmid DNA.

30 **Purification of DNA for microinjection.**

50-100µg of DNA is digested with Not I to separate the cVO14 cosmid insert from vector DNA. A salt gradient may be used as described by Dillon *et al.* (1993) to

purify the 44kb fragment. A gradient former is used to pour a gradient ranging in NaCl concentration from 5-25%. The digested DNA is applied to the top of the gradient which is then centrifuged at 5.5 hours at 37,000 rpm. The solution is then removed in 500µl aliquots which are examined by electrophoresis. Fractions containing the fragment to be
5 microinjected are pooled and ethanol precipitated. The pellet is dissolved in microinjection buffer (10mM Tris.HCl, pH 7.5, 0.1mM EDTA). DNA may be purified further using an Elutip column (Schleicher and Schuell) according to manufacturers instructions. cVO14 DNA at a concentration of 1-10ng/µl, typically 2ng/µl, is used to generate transgenic rats.

10

Superovulation, microinjection and embryo transfers.

40 day old prepubertal female Wistar rats are given intraperitoneal injections of 30 IU pregnant mare's serum (Folligon, Intervet Laboratories Ltd) between 9 and 11 o'clock on day -3. The same rats are injected i.p. at midday on day -1 with 22.5 IU human
15 chorionic gonadotrophin (Chorulon, Intervet Laboratories Ltd) and placed in a cage with a stud male of the same strain. On day 1, females are killed and their oviducts removed and placed in M2 media (Hogan *et al.*, 1986). The oviducts are dissected to release the eggs which are subsequently placed in M2 media with 0.5mg/ml hyaluronidase (Sigma) in order to remove the cumulus cells surrounding the eggs. After 5 minutes the eggs are
20 removed from the hyaluronidase solution, washed thoroughly in M2 and placed in the unbuffered M16 (Hogan *et al.*, 1986) in a 37 °C incubator supplemented with 5% CO₂. After 2 hours of incubation the male pronuclei of the eggs are microinjected using standard procedures and equipment (Hogan *et al.*, 1986). Microinjected eggs are incubated overnight at 37 °C.

25

The following day (day 2), eggs which have divided to the two-cell stage are washed in M2 media and transferred into the oviducts of pseudo-pregnant adult Wistar rats which have been mated with vasectomized male rats the previous night. The surgery is performed under halothane anaesthetic, with between 15 and 20 eggs being transferred
30 into each infundibulum. Resulting litters are tail-clipped at 2 weeks of age. The tails are used for DNA preparation as described above and analysed by Southern blotting for animals containing transgenes also as described above

RNA preparation.

RNA may be prepared from rat tissue by the acid guanidium thiocyanate-phenol-chloroform extraction method (Chomczynski *et al.*, 1987). Briefly, tissue is homogenised
5 in 500µl 4M guanidium thiocyanate, 25mM Sodium citrate (pH 7.0), 0.5% (w/v) sodium
N-lauroylsarcosine, 100mM 2-mercaptoethanol prior to the addition of 33µl 3M sodium
acetate (pH 4.1), 500µl phenol and 100µl chloroform. The mixture is vortexed and placed
on ice for 15 minutes before centrifugation at 12,000rpm for 10 minutes at 4 °C. The
aqueous fraction is decanted into a fresh tube and precipitated with isopropanol.

10

In vitro transcription.

A plasmid containing a T7 polymerase promoter 5' to the inserted sequence is
linearised with a restriction enzyme which cuts at the 3' end of the insert. Transcripts are
then obtained of the subcloned fragment using a T7 transcription kit (Boehringer
15 Mannheim) according to the manufacturer's instructions

DNA sequencing.

Sequencing of DNA plasmid subclones may be performed manually with the
Sequenase version 2.0 sequencing kit (United States Biochemicals) which employs the
20 chain-termination method (Sanger *et al.* 1977), or by automated sequencing using an ABI
Prism DNA Sequencing Kit and 377 DNA Sequencer (Perkin Elmer Applied Biosystems)
according to manufacturer's instructions.

Reverse Transcription of RNA.

25 RNA may be converted to cDNA using Superscript II reverse transcriptase
(GibcoBRL) according to the manufacturer's instructions, in combination with either an
oligo dT primer or another specific primer complementary to the RNA sequence of
interest.

30 Polymerase Chain Reaction amplification of DNA.

The polymerase chain reaction (PCR) may be used to amplify fragments of DNA
using 50µl of a reaction mix which contains 10mM Tris, pH8.3, 20mM KCl, 0.2mM

dNTPs, 200nM primers, 50-250ng template DNA, 2.5units Amplitaq DNA polymerase and 1-3mM MgCl₂ (the optimal conditions for each amplification are determined empirically). Conditions vary for each template target, but a typical amplification might be to place the reaction mix in a thermal cycler (MJ Research Inc.), denature for 2 minutes and then subject the reaction to 34 cycles of 94 °C for 1 minute, 58 °C for 1 minute and 72 °C for 1-5 minutes, depending on the length of the expected product.

Cloning of PCR products.

Products generated by PCR may be cloned using a TOPO TA Cloning kit (Invitrogen) according to the manufacturer's instructions.

Nuclease Protection assay.

Riboprobe transcripts incorporating ³²P may be generated by transcribing approximately 250ng of DNA fragment in Transcription Optimised buffer (Promega), 500mM ATP, GTP and CTP, 20 units Rnasin RNA inhibitor (Promega), 20mM UTP, 100μCi α-³²P-UTP (Amersham) and 20 units of the appropriate RNA polymerase (Promega) at 37 °C for 90 minutes. After treatment of the reaction with 2 units of DNase I (Promega) at 37 °C for 20 minutes, the reaction is denatured and purified by polyacrylamide electrophoresis, followed by excision of the labelled RNA and elution in 350μl of Elution buffer (Ambion) for 2 hours at 37 °C. Nuclease protection may be performed essentially as described by Lee and Costlow (1987) using ³²P-labelled riboprobe with 1-20μg total RNA and the RPAII Ribonuclease Protection Kit (Ambion) according to the manufacturer's instructions.

In Situ hybridisation.

Sense and anti-sense riboprobe transcripts may be generated using an SP6/T7 transcription kit (Boehringer) with ³⁵S-UTP (NEN Research Products) according to the manufacturer's instructions. *In situ* hybridisations may be performed as described in Bennett *et al.* (1995). Autoradiographs are analysed densitometrically, from a light box using a video camera linked to a Power Macintosh 7600/132 running the programme NIH Image version 1.61.

Immunocytochemistry

Human growth hormone (hGH) may be localised in pituitary and brain sections using a modified avidin-biotin complex immunocytochemistry technique (Bourne *et al.*, 1984). Tissue is collected and fixed in 4% paraformaldehyde for 24 hours. Tissues may be
5 stored at 4 °C in 70% ethanol before embedding in paraffin wax and sectioning. Tissue sections (6µm) are dewaxed in Histoclear (National diagnostics) and rehydrated by sequential 20 sec washes of 100%, 70 % and 30 % ethanol followed by a 1 min wash in distilled H₂O. Endogenous peroxidase activity is inhibited by a 30 min incubation in 3% (v/v) hydrogen peroxidase in methanol. Sections are then washed in distilled H₂O for 1
10 min before being treated with 0.1% (w/v) trypsin (Sigma) for 15 min at 37 °C followed by 0.5% (v/v) Triton X-100 (Sigma) for 15 mins. After two 5 min washes of distilled H₂O and 0.05M Tris buffered saline (pH 7.6), 0.15M NaCl (TBS) the sections are incubated with 20% (v/v) normal rabbit serum (DAKO) with 5% (w/v) BSA for 30 mins in order to reduce non-specific background staining. The sections are then incubated overnight in a
15 humidity chamber at 4 °C with an antibody specific for hGH, such as sheep anti-hGH primary antibody (1:30,000) (Scottish Antibody Production Unit).

Following two washes in TBS, sections are incubated with biotinylated rabbit anti-goat serum (DAKO, 1:200) for 30 mins. The sections are again washed in TBS and incubated
20 for 30 mins with avidin complexed to biotinylated horse radish peroxidase (DAKO). Human GH immunoreactivity is then visualised by development using 3,3-Diaminobenzidine tetrachloride/hydrogen peroxide (DAB) (4mg/10ml in 0.05M Tris buffer pH 7.6) containing 3% (v/v) H₂O₂. This reaction is quenched in distilled H₂O prior to counterstaining with Gill's haematoxylin (BDH) and coverslipped for microscopic
25 examination.

Radioimmunoassays

Tissue samples are homogenised in varying volumes of phosphate buffered saline with either glass homogenisers (for volumes up to 1ml) or a Kinematica Polytron PT 3000
30 homogeniser (for larger volumes). The same polyclonal sheep anti-hGH antibody (Scottish Antibody Production Unit) may be used to measure hGH in tissue extracts by RIA as described in Fairhall *et al.* (1992) using recombinant hGH as standard. AVP may

be measured by RIA as described in Horn, Robinson & Fink (1985). Rat GH may be measured as in Charton *et al.*, (1988). Bovine neurophysin may be measured by RIA using a specific antiserum that does not recognise rat neurophysins (Gordon Weeks, 1987). Rat leptin and rat insulin may be measured by specific RIAs using kits from Linco Research Inc, following the manufacturer's instructions. Corticosterone may be measured by a double antibody RIA kit obtained from ICN Biomedicals. Cholesterol and triglycerides in blood samples may be measured using kits obtained from Sigma Diagnostics ('Cholesterol 20' and 'Triglycerides, UV'). Plasma glucose values may be measured using a Beckman glucose analyser.

10

EXAMPLE 1

ISOLATION OF COSMID DNA, CONSTRUCTION OF TRANSGENE COSMIDS AND GENERATION OF TRANSGENIC RATS

15 Isolation of cosmid DNA

Since the DNA sequences of the rat AVP and OT genes and their orientation and structural relationship to each other in the rat genome are known (Ivell & Richter, 1984a; Mohr *et al.* 1988; Schmitz *et al.*, 1991) the size of restriction fragments which should be detected with cDNA probes for these genes can be predicted. Colonies bearing rat DNA which contained fragments hybridising to these OT and AVP probes in the same areas in duplicate filters from the cosmid screening are aligned with the original bacterial plates. These colonies are picked and grown and DNA prepared from them, digested with Hind III, run on agarose gels, Southern blotted and hybridised to the same OT and AVP probes again. Three positive colonies are chosen for further analysis because their differing restriction fragment patterns indicated that they spanned different regions of the rat AVP/OT locus. From Southern blotting of restriction digests using probes against the first exon of each gene and the vector, they are found to span a total of 44kb, including both genes, the 11kb intergenic region, 8kb of AVP 5' flanking sequence and 24kb of OT 5' flanking sequence. These three overlapping cosmids are designated cVO1, 2 and 3. An overall schematic map of this region, indicating the location and orientation of AVP and OT genes, the location of some important restrictions sites, and areas of sequence known or subsequently determined and disclosed here, is shown in Figure 1

30

To facilitate further restriction mapping of the 5' flanking sequence of the rat OT gene, 8kb and 14kb Sma I fragments and an 8.5kb Kpn I fragment are subcloned into cloning vectors and subjected to further restriction mapping. Smaller fragments of the OT and AVP genes are also subcloned into pUC 19 derived plasmid vectors and used to remove restriction enzyme sites and to insert the reporter genes into the rat OT and rat AVP loci. Oligonucleotide linkers containing sequences for unique restriction sites are also inserted in the 5' untranslated regions of the two genes to allow for future modifications of this construct.

10

The subcloning strategy used for the AVP locus is outlined in Figure 2. Essentially, the aim is to insert a genomic hGH reporter fragment in a unique cloning site introduced into the 5' untranslated region of rat AVP gene. Swanson *et al.* (1985) had previously shown that hGH reporter transcripts, when fortuitously expressed, may be expressed and translated efficiently in these neuronal cells types. Furthermore, hGH nucleotide sequences can be differentiated from rat GH sequences by specific nucleotide probes (Seeburg *et al.*, 1977; Roksam & Rougeon, 1979) and the protein can be differentiated from rat GH by specific antisera (Appendix 1). An Mlu I linker is initially inserted into a smaller subclone of the rat AVP gene, replacing the Dra III site in the 5' untranslated region. A genomic fragment of the human GH structural gene (Roksam & Rougeon, 1979) is then inserted as an Mlu I-linkered fragment spanning from the 5' untranslated region of the hGH gene to a region 3' of the last exon and containing all 5 exons and 4 introns of hGH. This AVP-hGH fragment is inserted as a 12.2 kb Cla I-Xho I fragment containing 450bp 5' and 8kb 3' of the transgene, with deletion of other Xho I restriction sites.

25

The subcloning strategy used for the OT locus is outlined in Figure 3. In this case the aim is to replace most of the rat OT structural gene with corresponding bovine sequences (Land *et al.*, 1983). The protein produced should function identically, but the bovine sequences would provide a 'silent' reporter since they could be differentiated from rat sequences (Mohr *et al.*, 1988) by specific nucleotide probes, and the protein differentiated from rat neurophysin by specific antisera. Rat neurophysin has previously been used as a

30

transgene reporter in mice (Belenky *et al.*, 1992). Due to the constraints of suitable restriction sites it is necessary to assemble a 5' construct of the hybrid rat/bovine OT gene (containing exon A and most of exon B) and a 3' construct (containing a small fragment of exon B and exon C) separately. These constructs are joined to produce the hybrid gene with the 5' and 3' flanking sequences being added in subsequent steps. A Sal I linker is also inserted immediately 5' to the translational start site of the bovine OT to provide a unique cloning site within the AVP/OT locus for future modification of the construct. The hybrid gene is inserted into the final construct as a 10.5 kb Mun I - Xho I fragment containing 7.8 kb of 5' and 1.7 kb of 3' flanking sequence, after deleting other restriction sites within the cosmid

Assembly of the final construct.

The pWE15 vector is modified to remove the unrequired SV2 neomycin gene. This reduced the vector size from 8.5kb to 4.2 kb and therefore increased the size of the insert that could be subcloned into the cosmids, which can efficiently package up to 52 kb (Wahl *et al.* 1987). Cla I, Mun I, Sal I and Mlu I restriction sites are also removed from the vector to permit subsequent cloning steps. The pWE15 cosmid vector has Not I sites flanking the insert. Restriction mapping of cVO1 revealed a Not I restriction site 13kb upstream from the rat OT gene (Figure 1) which would also be digested if Not I is used to remove the insert. Therefore, a 4.6kb Aat II-Sca I fragment containing this site is subcloned, the Not I site deleted, and the fragments ligated and replaced into the construct. Digestion of the ligation product with Not I confirmed that this site had been destroyed.

25 The modified OT and AVP gene fragments are inserted into cVO3, followed by addition of the 5' region present in cVO1 but not in cVO3 using an Aat II fragment. This adds all except 1.1kb of the extreme 5' end of cVO1 and contains a Not I linker at the extreme 5' end.

30 The final construct, termed cVO14, spans 44kb and includes 8kb 5' of rat AVP, 24kb 5' of rat OT and 11kb of intergenic sequence. The construct has reporter gene hGH sequences inserted into the 5' untranslated region of the rat AVP gene and parts of the

bovine OT gene sequences substituted for equivalent rat OT gene sequences. The final cVO14 construct is illustrated diagrammatically in Figure 4.

5 **Generation of transgenic rats bearing the cVO14 construct.**

 The 44 kb Not I insert is released from cVO14 by Not I digestion, purified on a salt gradient and microinjected into fertilised rat oocytes. These embryos are transferred into pseudopregnant mothers and the offspring are analysed for the presence of the transgenes. Genomic DNA prepared from tail biopsies of these pups is digested with Bgl
10 II, Southern blotted and hybridised with a radiolabelled genomic hGH probe that should identify 2 predicted fragments of 0.9 and 2.1 kB from transgene DNA. This probe does not detect endogenous rat GH sequences. Of 102 pups the hGH transgene is present in the DNA of only 3 pups, termed JP 17, JP 19 and JP 59. JP 19 dies at 11 days of age, and is not analysed further.

15

Other samples of DNA from the two remaining rats is digested with Pst I, Southern blotted and hybridised with a radiolabelled probe that should identify two predicted fragments of 0.9 and 1.6kB from the hybrid rat/bovine transgene sequence, and a single
20 2.5kB fragment from the endogenous OT gene. Both JP17 and JP59 rats are also found to contain this hybrid gene, as well as the endogenous gene, whilst only the endogenous fragment is visualised in DNA from non-transgenic rats.

DNA from JP 17 and JP 59 rats is also Southern blotted and probed with radiolabelled DNA fragments corresponding to the ends of the cVO14 construct, which confirmed that
25 whole copies of the microinjected fragment are present in both rats. The copy number of the transgenes is estimated by Southern blotting of Hind III fragments and hybridisation with a probe for the rat AVP gene sequences, which recognised a 3.4 kb fragment corresponding to the endogenous rat AVP gene and a 5.2 kb fragment which represents the transgene with its hGH reporter gene insertion. Assuming equal affinity to
30 endogenous or transgene sequences, phosphorimaging these blots suggested that the JP17 rats contained at least 4 copies of cVO14 whereas JP59 rats had a single copy. The copy

number and restriction pattern of the transgenes remained consistent through successive generations of breeding, suggestive of a single site of chromosomal integration.

Further analysis of DNA suggested that the insert contained concatamers of cVO14 in JP17 but not in JP59, and that one concatamer pair contained a truncation which removes a fragment of approximately 1kb between 8kb and 7kb 5' of rat AVP. Restriction mapping and sequence analysis of the cosmid ends enabled us to design PCR primers (PL216 (SEQ. ID. No. 10 and PL210 (SEQ. ID. No. 9)) that uniquely identify DNA from JP17 rats bearing this insert, and distinguishes them both from non-transgenic littermates, and from JP59 rats bearing a single copy of this insert.

Establishing colonies of transgenic rats.

The founder JP17 rat is a male. He sires only single litter of rats at 6 months of age although constantly caged with fertile females. This litter contains both male and female rats bearing the transgene, indicating that the integration has occurred onto an autosomal chromosome. No further litters are sired by male progeny. Litters bred from transgenic JP17 female progeny show an approximate 1:1 ratio of transgenic to non-transgenic rats (46 transgenic versus 54 non-transgenic in the first 100 pups) suggesting that the transgene does not have a detrimental effect on embryonic viability.

20

The founder of the JP 59 line of rats is female and bred normally (the ratio of transgenic to non-transgenic pups is approximately 1:1 (47 transgenic verses 53 non-transgenic in the first 100 pups). This single copy integrant is also present on an autosomal chromosome since male JP59 rats of this line sire transgenic progeny of both sexes.

25

EXAMPLE 2

ANALYSIS OF EXPRESSION OF EXPECTED TRANSGENE PRODUCTS

Human growth hormone

30

The expression of hGH from cVO14 is investigated in both JP59 and JP17 rats. Immunocytochemistry shows expression of hGH protein in hypothalamic magnocellular paraventricular (PVN) and supraoptic nuclei (SON). Human GH immunoreactivity is also

transported via axons passing through the internal zone of the median eminence and present in the axon terminals in the posterior pituitary.

In situ hybridisation confirms strong expression of hGH in the PVN and SON of transgenic rats of both JP59 and JP17 lines. hGH transcripts are also detected in other sites of AVP expression in the CNS in JP 17 transgenic rats, such as the medial amygdaloid nucleus and the habenula (Buijs, 1987; Caffé *et al.*, 1987; Urban *et al.*, 1990). Double *in situ* hybridisation analysis or immunocytochemical analysis confirms that hGH expression is localised in AVP neurones, and not in OT neurones. In independent studies, RT-PCR analysis detects hGH transcripts in hypothalami and pituitaries from both lines, and also detected transcripts in the pancreas and also faintly, in adrenals of JP 17 rats, but not in other tissues tested. These findings are in accordance with previous reports of extrapituitary expression of the endogenous AVP gene in these tissues.

Radioimmunoassay for hGH confirms the presence of significant quantities of hGH in posterior pituitary extracts from both JP59 and JP17 animals, with larger amounts in JP17 line consistent with their higher copy number. Small amounts of hGH immunoreactivity are also found in the pancreas (0.016 ± 0.0075 ng/mg of tissue, $n=3$) of 20 week old JP 17 male rats, though this represents $< 0.1\%$ of the amounts of hGH found in the posterior pituitary extracts of the same rats ($168\text{ng} \pm 16$ ng/mg, $n=5$). Thymus, heart, kidney, fat, liver, ovary, uterus, testis, lung, cortex, cerebellum, spleen and adrenals all had undetectable levels of this protein ($<0.0004\text{ng}$ of hGH/mg of tissue).

If the hGH transgene is correctly expressed, then stimuli for increased AVP synthesis and release should increase hypothalamic expression of the hGH transgene and decrease pituitary stores of hGH. Chronic osmotic stimulation has been shown to regulate the expression of the AVP gene (Lightman *et al.*, 1987; Murphy *et al.*, 1990) and cause a release of AVP from the posterior pituitary (Fitzsimmons *et al.*, 1994). The stimulus of salt-loading has previously been used to detect whether the DNA regulatory regions responsible for physiological regulation of the rat AVP gene are present within microinjected constructs (Zeng *et al.*, 1994b; Waller *et al.*, 1996). Groups of non-transgenic or transgenic JP59 or JP17 male rats given 2% NaCl w/v in their drinking

water for 72 hours both show a marked increase in hGH expression in PVN and SON. Furthermore, posterior pituitary hGH content fall significantly in such salt-loaded animals in parallel with the fall in AVP content. Samples taken from JP17 rats confirm that hGH is secreted, and can be detected in plasma by RIA (1.3 ± 0.09 ng/ml)

5

The effects of transgenic expression of hGH to reduce rat GH by feedback have been documented earlier in other transgenic rats (Flavell *et al.*, 1996). Therefore, rat GH content of the pituitaries of JP 17 rats and non-transgenic littermates are measured by RIA. Rat GH is significantly reduced in both the male and female JP 17 transgenics in comparison to the non-transgenic controls at 23, 77 and 140 days of age. At 140 days, the mean pituitary rat GH content of male JP 17 rats is 34.2% of that of the age-matched non-transgenics. The pituitary rat GH content is less affected in the female JP 17 rats (57.4% of the mean rat GH content of the non-transgenics, $p < 0.02$).

15 The size of the anterior pituitaries also suggests that there is a reduction in their cell number as JP 17 male rats at 140 days have significantly smaller anterior pituitaries than non-transgenic controls (4.6 ± 0.1 mg for JP 17 males versus 8.2 ± 0.5 mg for wild-type rats, $p < 0.002$, $n=6$). The pituitaries of JP 59 males and females do not show a reduction in rat GH content or size ($p > 0.55$).

20

Bovine neurophysin

No bovine OT-NP protein can be detected in posterior pituitary extracts (<10 pg per pituitary) from JP17 rats, using a specific RIA that distinguishes bovine neurophysin from rat neurophysins (Gordon-Weeks, 1987). An RT-PCR assay for bovine OT-NP transcripts is applied to hypothalamic extracts of adult males or of lactating female rats culled within 24 hours of littering, from both lines. The latter animals are chosen since they should show higher levels of endogenous OT expression (Van Tol *et al.*, 1988). PCR is performed on cDNA generated by reverse transcribing RNA from various tissues of both JP 17 and JP 59 lines (hypothalamus, pituitary, pancreas, ovary, heart, lung, muscle, thymus, cerebellum, uterus, testis, spleen, kidney, adrenals, liver, cortex). Additional reactions with primers for β -actin and hGH are also included. The reactions of the rat/bovine OT primers with the cVO14 construct and in vitro transcribed rat/bovine

30

OT RNA both yielded the correct size fragment (767bp), but no transcripts from the rat/bovine OT transgene are detected in any tissue. We conclude that the rat/bovine OT portion of the cVO14 construct is not detectably expressed in JP 17 or JP 59 transgenic animals.

5

The lack of expression of the rat/bovine OT transgene may indicate that additional sequences lacking from cVO14 are required to achieve appropriate OT expression in addition to expression from the AVP locus. Other possibilities are that alterations introduced into the OT locus prevent expression. This could be in the coding regions of the hybrid rat/bovine OT cassette. Another possibility is the introduction of the Sal I linker 5' of the OT gene. A further possibility is the presence of a base change in the region immediately 3' of the TATA-box which is discovered upon sequencing this region of cVO14 (Figure 5).

10

15

EXAMPLE 3

DISCOVERY AND ANALYSIS OF 5'OT-EST AND 5'OT-EST-XDEL

20

25

30

cVO14 is noted to contain a CpG island 13kB upstream of OT. Sequencing of 3.3kB of this region of the cosmid reveals a potential novel gene. Comparison with EST sequences in the public databases revealed partial matches to sequences from rat, human and mouse origin. The GenBank accession numbers for such ESTs include: H31114; H31115; AA955566; AA850004; AA104183; AA080247; AA245389; AA242211; AA421310; AA505752; AA421393. Such searches also reveal a partial match to a human genomic sequence GenBank Accession number AF036329. From comparisons with these sequences and the rat genomic DNA sequence disclosed herein, it is predicted that the novel rat gene in cVO14 contains four open reading frames, termed w, x, y, z. This gene is termed 5'OT-EST. The genomic DNA sequence and predicted exon structure is disclosed in SEQ. ID. No. 16. The gene predicts an open reading frame (SEQ. ID. No. 1) and a protein of 200 amino acids, termed 5'OT-EST, whose structure is disclosed in SEQ. ID. No. 2. Comparisons with EST sequences from human and mouse sources and alignment with full length sequences from rat DNA enable the prediction of homologous cDNA and protein sequences in these species and these are disclosed in SEQ. ID. No. 3.

and 4 and SEQ. ID. No. 5 and 6 respectively. The protein sequences predicted from these predicted cDNAs are highly homologous, as shown in Figure 6.

5 A Not I restriction site is identified approximately 13kB upstream of OT in cVO3. As described in Appendix 2, this site is deliberately destroyed during the construction and assembly of cVO14 in the pWE15 cosmid vector, as the construct required Not I sites only at the ends of the insert. However, sequence analysis of cVO14 reveals that the Not-1 site lies in 5'*OT-EST*, more precisely, in exon w of 5'*OT-EST*. Furthermore, this sequence analysis reveals that in addition to destroying the Not I site, the procedure used
10 (digestion, filling in and religation) also resulted in an additional unpredicted deletion of 412bp. This deletion includes all of the sequences recognised as exon x as defined herein. The mutated form of this gene, lacking sequences including those for exon x, is therefore termed herein 5'*OT-EST-xdel* for the purposes of this application. Its sequence, and the structures of the predicted exons from this form of the gene are disclosed in SEQ. ID. Nos.
15 7 and 8. The presence of 5'*OT-EST-xdel* in the genome of JP17 and JP59 rats is confirmed by the generation of the predicted shorter product upon amplification of genomic DNA from these animals by PCR with primers PL266 (5'TCATGTTGCGGGCTTTGAAC) and PL271 (5'TCTTTCAGTTGCACCCAAGC) which flank the deletion (see SEQ. ID. Nos. 11 and 12 respectively).

20 The form of 5'*OT-EST* that is incorporated in both JP17 and JP59 in 4 or 1 copies, respectively, is mutated from the wild type sequence. 5'*OT-EST-xdel* would be predicted to give rise to an altered mRNA, which if translated would produce a truncated protein product with an additional novel amino acid sequence. The predicted sequence of this
25 novel product, termed herein 5'*OT-EST-xdel* is disclosed in SEQ. ID. No. 8. Comparison with the aligned predicted protein sequences of 5'*OT-EST* in normal rats and in other species predicts that the protein translated from this RNA would contain an altered exon w, with a novel C-terminal peptide sequence (shown in lower case beginning at the arrow in Figure 6) predicted to arise by translation of DNA sequences normally present as part
30 of an intron in 5'*OT-EST*. Searches in the protein databases in the public domain do not find any significant matches of this mutated protein sequence to known sequences.

To demonstrate that both the endogenous and truncated forms of *5'OT-EST* are transcribed in JP59 and JP17 rats, PCR primers are designed which can distinguish between these gene products. The sequence of these primers is given in SEQ. ID. No. 11, 12 and 13. RT-PCR using these primers confirms the presence of transcripts from the endogenous form of *5'OT-EST* in testicular RNA extracts from JP17, JP59 and wild-type rats, but the presence of a transcript with the 412bp deletion only in such tissue extracts from JP17 and JP59 rats. Sequencing of amplification products generated by PCR with primers PL266 (SEQ. ID. No. 11) and PL273 (SEQ. ID. No. 13) from wild type and JP17 rats confirms this region of the sequence of the endogenous rat transcript as well as the truncated *5'OT-EST-xdel* sequence disclosed in SEQ. ID. No. 8. Extracts of a rat adrenal medullary cell line (PC12 cells) also contain an RNA product of *5'OT-EST* of the expected size.

Identification and sequencing of *5'OT-EST* and *5'OT-EST-xdel* enables the design of probes to carry out *in situ* hybridisation and RNase protection analysis for the products of these genes on normal and JP17 rat tissue extracts. *In situ* hybridisation with probes complementary to exons w or z (more specifically, corresponding to bases 1020-1167 and 2229-2451 of Fig 4.1 respectively) on hypothalamic sections from wild-type or transgenic animals, revealed a highly specific expression in magnocellular SON and PVN. No other specific expression in different brain regions is observed at this level of detection. This is an unexpected finding, which is repeatedly confirmed. *5'OT-EST* is a novel member of the AVP/OT locus and is expressed in the same hypothalamic magnocellular neurones. Similar patterns of expression are seen with both probes and no differences in tissue distribution of hybridisation signal are seen between wild-type or JP17 tissues. Further *in situ* hybridisation analysis on a wide variety of tissues reveals strong expression in the testis consistent in distribution with tubular or Sertoli cell expression. Sparse expression is also seen in other tissues, including lung, spleen, intestinal smooth muscle and adrenal gland.

From the sequence information, it is further possible to design probes for *in situ* hybridisation analysis that distinguish completely between the forms of mRNA produced from *5'OT-EST* and *5'OT-EST-xdel*. More specifically, oligonucleotide probes directed

against transcripts containing exon x would be predicted to detect *5'OT-EST* but not *5'OT-EST -xdel* transcripts, whilst probes directed against the intron sequence in *5'OT-EST* that immediately follows the truncation in *5'OT-EST -xdel* detect transcripts containing this sequence, that code for the truncated product in extracts from rats
5 expressing *5'OT-EST -xdel* transcripts (such as JP17 and JP59 rats) but not from non-transgenic rats. Examples of such probes are given in SEQ. ID. Nos. 14 and 15.

An oligonucleotide probe of the sequence depicted in SEQ. ID. No. 15 (specific for the truncated sequence) is used for in situ hybridization and confirms transgene expression
10 specifically in PVN and SON in JP17 rats, whereas no signal is observed in PVN or SON sections from non-transgenic rats, hybridized at the same time with this probe.

Nuclease protection analysis may also be performed using a riboprobe to exon w as described above. From the sequence we disclose herein, this probe would be predicted to
15 protect 147bp and 94bp bands from transcripts from *5'OT-EST* and *5'OT-EST - xdel* respectively. Using such a probe to analyse testicular RNA extracts confirmed that the full length transcript is present in both transgenic and non-transgenic animals and that the truncated product is present in JP17 and JP59 extracts in a level consistent with the copy number of the cVO14 insertion, that the truncated transcript is indeed absent from non-
20 transgenic testis extracts. The full length product is present in control extracts of PC12 cells.

The gene termed *5'OT-EST -xdel* present in cVO14 in both JP59 and JP17 rats is transcribed in several tissues in JP17 rats, specifically in hypothalamic cells and in
25 testicular cells, and that the sequence of the truncated transcripts if translated, would give rise to a protein product that is severely truncated with respect to the normal gene product and would contain additional novel peptide sequence.

EXAMPLE 4**PHENOTYPIC ANALYSIS OF JP17 TRANSGENIC (SLOB) RATS****Growth measurements**

5 JP 17 transgenic rats of both sexes and non-transgenic littermates are weighed at regular intervals. Male JP 17 rats show a slight but significant reduction in their body weight up to 120 days of age ($p < 0.01$). This juvenile growth retardation is not seen in females of this line or the rats of either sex of the JP 59 line whose body weights are not significantly different to those of the non-transgenic groups ($p > 0.7$). This effect
10 disappears with time. At 140 days, the weight difference between JP 17 and non-transgenic male rats is no longer significant. Some organs of 140 day old rats are dissected and weighed. Heart weights do not differ significantly ($p < 0.14$) but the weights of the kidneys ($0.99 \pm 0.03\text{g}$ in JP 17 rats versus $1.24 \pm 0.05\text{g}$ in non-transgenic rats, $p < 0.001$), liver ($11.11 \pm 0.29\text{g}$ in JP 17 rats versus $14.40 \pm 0.32\text{g}$ in non-transgenic rats, $p < 0.001$) and spleen ($0.66 \pm 0.02\text{g}$ in JP 17 rats versus $0.84 \pm 0.03\text{g}$ in non-transgenic rats, $p < 0.001$) differs in weight ($n=6$ in all groups). Disproportionate growth is well
15 known in transgenic animals expressing hGH (e.g. Shea *et al.*, 1987)

Body weight measurements in ageing JP 17 rats.

20 After about 140 days, the group of JP 17 transgenic male rats gain weight more rapidly than their non-transgenic littermates (Δ weight between 200 and 420 days $356.5 \pm 57.419\text{g}$ for JP 17 males versus 182.50 ± 7.554 for non-transgenic males, $p < 0.03$, $n=5$, Figure 5.1). Female JP 17 transgenic rats show only a slight increase in weight gain when compared to non-transgenic littermates (Δ weight between 280 and 480 days $111.8 \pm 8.2\text{g}$
25 for JP 17 females versus $88 \pm 5.1\text{g}$ for non-transgenic females, $p < 0.04$, $n=6$). This is illustrated in Figure 8, which clearly shows the sexually dimorphic weight gains in these animals. No significant increased weight gain is observed in either sex of transgenic JP 59 rats compared with non-transgenic JP59 rats. At one year, the weights of the kidneys and liver of JP 17 male rats have reached a value that is not significantly different than that of
30 the non-transgenic rats ($n=6$ in both groups) ($p < 0.08$ for kidneys and livers), but the spleens remain lighter ($1.03 \pm 0.04\text{g}$ versus $1.225 \pm 0.05\text{g}$ in wild-type rats, $p < 0.01$).

These organs in transgenic JP 59 rats show no variation from their non-transgenic littermate controls ($p>0.43$).

Body length, width and fat-pad measurements.

5 Measurements of the body lengths (nose-anus) and the width across the pelvic area of anaesthetised male JP17 and non-transgenic rats are taken (Figure 9). At 20 weeks of age male JP 17 transgenic rats are shorter than their littermate non-transgenic controls with an increased width across the pelvic area. At 52 weeks, the difference in nose-anus length is no longer significant but the girth of the transgenic JP17 rats has increased
10 greatly whereas the non-transgenic rats only exhibit a moderate increase in girth. This late-onset increase in the body weight/length ratio is shown in Figure 10.

A comparison of the body proportions of a live SLOB rats and non-transgenic littermates shows a marked increase in abdominal fat. This is also obvious when individual peri-renal fat pads are compared. To evaluate this abdominal distribution of extra fat in SLOB
15 rats, peri-renal and testicular fat pads are dissected and weighed from matched groups of male JP 17 and non-transgenic littermates of 77, 140 and 365 days of age, and the results are shown in Figure 11. The peri-renal fat pads of the transgenic rats are markedly increased in weight at both 140 days and 365 days when their mean weight is almost five
20 times that of the non-transgenic animals. The testicular fat, however, did not show a comparable increase. Although testicular fat pad weights are marginally larger than those of the non-transgenic rats at 140 days ($p<0.05$), no further significant increase occurs during the period of a large accretion in peri-renal fat, and there is no difference in testicular fat pad weights at 365 days between SLOB rats and their non-transgenic
25 littermates, despite their much larger body weight, and evident gross visceral obesity.

In other matched groups of 1 year old SLOB rats and non-transgenic littermates of both sexes, plasma cholesterol, triglycerides, glucose, insulin, leptin and corticosterone are measured in blood samples taken when the animals are killed, and the results are
30 summarised in Figure 12. Plasma triglycerides are modestly but significantly elevated in SLOB males compared to non-transgenic males. There are no differences in plasma triglycerides in females. Cholesterol levels are no different between the groups. Plasma

glucose and insulin values are also in the normal range and did not differ between the groups, suggesting that the obesity is not secondary to diabetes or insulin resistance. Plasma corticosterone is also in the normal range in all groups of rats. Notably however, the plasma leptin levels are elevated significantly in both male and female SLOB rats compared with their non-transgenic littermates, and are almost two-fold higher in SLOB males than in SLOB females. Leptin receptor transcript isoforms are also expressed in normal amounts in the hypothalamus, piriform cortex and choroid plexus. These increases would be expected given their increased body fat, but prompted a study of their food intake.

10

A further group of 5, 11-month old SLOB rats and 5 non-transgenic rats are housed singly in metabolic cages for 14 days, and after a period of acclimatisation to single housing, food intake is measured over the last four days of the experiment. There is no significant difference in food consumption between the two groups (SLOB rats 23.4 ± 1.1 g/day vs. 23.5 \pm 1.8 g/day in the non-transgenic males, Mean \pm S.D.).

15

Although the SLOB phenotype, as demonstrated by the forgoing, has a striking late-onset feature, the phenotype is latent at a younger age and can be induced by increasing the levels of fat in the diet. This is demonstrated by observing the phenotypic differences resulting from feeding two groups of 100 day old transgenic and normal littermates either regular rat chow, which has a fat content of 4%, or a high fat diet having a fat content of 30% over a 27 day period.

20

The rats fed a normal diet show no significant difference in weight gain between transgenic and non-transgenic littermates. However, in the case of the rats fed on a 30% fat diet, the transgenic animals gain twice the weight of their non-transgenic littermates (see Figure 13). Controls in dwarf rats show that the obese phenotype is not due to growth hormone deficiency.

25

Plasma leptin levels are measured at sacrifice. These are found to be higher in transgenic animals, and rise in both transgenic and non-transgenic rats fed on a 30% fat diet.

30

Moreover, the increase in dietary fat is associated with a significantly reduced food intake in normal rats, but not in SLOB rats, despite their higher leptin levels.

Induction of obesity in ovariectomised female rats.

5 Four groups of female rats are studied (see Figure 14). Sham-operated transgenic female SLOB rats are lighter than non-transgenic sham-operated female littermates at 100 days, but gain the same amount of weight in the following 11 week period ($\Delta\text{wt } 45.5 \pm 5.3\text{g}$, vs $\Delta\text{wt } 48.4 \pm 3.8\text{g}$). In rats ovariectomised under anaesthesia, both groups show an increase in weight gain; this increase is much higher in SLOB rats than in non-transgenic
10 littermates ($\Delta\text{wt } 128 \pm 7.7\text{g}$ vs $89 \pm 4.3\text{g}$, $P < 0.001$). Some animals from each group are killed 18 weeks post ovariectomy and their supra-renal fat pads dissected and weighed. The fat pad weight is much larger after ovariectomy in SLOB rats (4.67 ± 0.61 vs $1.37 \pm 0.39\text{g}$ in ovariectomised versus sham-ovariectomised SLOB females, $P < 0.01$), than in nontransgenic rats ($2.33 \pm 0.94\text{g}$ vs $1.0 \pm 0.01\text{g}$ in ovariectomised vs sham-
15 ovariectomised nontransgenic littermates).

Fertility of the JP 17 male rats.

Twelve JP 17 and twelve non-transgenic young adult males are each housed with two 12-week old normal females, for several consecutive days. The female rats are
20 examined every morning for evidence of copulation, either in the form of a vaginal plug or sperm in vaginal smears, and are observed for a sufficient amount of time to allow any litters conceived during this time to be born. No litters are sired in this time by JP 17 males, whereas 11 of the 12 females housed with wild-type males produced litters.

25 The immediate cause of infertility in male SLOB rats is unknown. The size and gross anatomy of their testes and seminal vesicles is normal, suggesting unaffected levels of gonadotrophins or androgens. Testicular size, sperm morphology, motility and testosterone levels all appear normal in SLOB rats. Treatment of a SLOB male rat with exogenous androgens did not improve fertility. One cause could be hGH, since infertility
30 is a common problem in GH transgenic animals (Yun *et al.*, 1987, Bartke *et al.*, 1988, Flavell *et al.*, 1996) and male transgenic animals expressing hGH have been reported to have a reduced frequency in the impregnation of females (Bartke *et al.*, 1992). However

female SLOB rats also express equivalent amounts of hGH and are not infertile. Furthermore, we found no evidence for hGH expression or of hGH protein in testes from SLOB rats.

- 5 In contrast, the expression of *5'OT-EST* in normal rats and the high level of a truncated RNA product from *5'OT-EST -xdel* in hypothalamus and in particular, the testis from SLOB male rats, and the lack of expression of either product in ovaries in SLOB females, leads us to conclude that the novel infertility and obesity phenotype more probably results from the presence of multiple copies of *5'OT-EST -xdel* in SLOB rats. A disruption of
- 10 testicular function by *5'OT-EST -xdel* and consequent infertility is part of, may partly contribute to, or exacerbate the degree of the male-preponderance of, the obesity phenotype of SLOB rats. A testicular disruption is not absolutely required however, since a mild visceral obesity can also be discerned in SLOB females.

15 **Longevity of SLOB male rats.**

The longevity of JP 17 also appears to differ to that of normal rats. Six male JP 17 rats and six wild-type rats are housed under constant conditions. After two years, all six JP 17 rats have died, five at between 10 and 14 months of age and the sixth at 21 months of age, whereas only a single wild-type rat has died at 13 months. The longevity of JP 17

20 females or JP 59 males or females has not been similarly investigated.

Comparison of phenotype with other rat obesity models

When comparing the phenotype in SLOB rats with other findings reported in the literature, the closest parallels are lines of transgenic rats expressing hGH driven by a mouse whey acidic protein promoter, (Ninomiya *et al.*, 1994; Ikeda *et al.*, 1994, 1995, 1997). Ikeda *et al.* (1995) described two lines of rats expressing high or low hGH levels in serum. Gigantism is observed in the high hGH-expressing line, but visceral obesity is also observed in the low-expressing line, associated with endogenous GH suppression. No sexual dimorphism is reported, and the obesity is associated with carbohydrate

25 metabolic disorders, hypertriglyceridaemia and insulin resistance. Ikeda *et al.* 1995 specifically concluded that the effect is due to differences in serum hGH levels affecting

30 carbohydrate metabolism. A later study in these rats (Ikeda *et al.* 1997) reported female

infertility and enlarged ovaries which further distinguishes this phenotype from that seen in SLOB rats.

In common with the rats reported by Ikeda *et al.* (1994), SLOB rats also show reduced rat
5 GH production and secretion. GH deficiency is associated with increased visceral fat in humans, but this can be alleviated by hGH treatment. However, isolated rat GH deficiency is an unlikely cause of obesity in SLOB rats as other lines of severely GH deficient dwarf rats (Charlton *et al.*, 1988) do not develop obesity when housed under identical conditions to SLOB rats. Obesity can be induced in such dwarf rats (as in
10 normal rats), when placed on high fat diets for prolonged periods though females are more susceptible than males (Clark *et al.*, 1996). A similar pituitary GH suppression is also evident in female SLOB rats but they do not develop the same massive abdominal obesity as males. Pituitary rat GH suppression is also seen in the non-obese JP59 rats of both sexes and in Tgr rats (Flavell *et al.*, 1996) which do not develop obesity.

15

The defects in other genetic models of obesity in the rat have recently clarified; examples of these include the Zucker fa/fa rat, the Koletsky (f) obese rat, the JLA/cp corpulent rat, and the OLETF rat, and their related sub strains (Iida *et al.*, 1996; Wu-peng *et al.*, 1997; Takaya *et al.*, 1996; Lee *et al.*, 1997; Kahle *et al.*, 1997;. None of these show the male
20 specificity, late onset or pattern of distribution of obesity seen in SLOB rats and they exhibit significant hyperglycaemia and insulin resistance, which again distinguishes them from SLOB rats. Male specificity, infertility, extremely late onset of obesity, a highly selective visceral accumulation of fat, but relatively normal metabolic profile, without insulin resistance, hyperphagia or hyperglycaemia distinguishes the dominant phenotype
25 in the SLOB rats from all other known models of obesity in the rat, including those with low endogenous rat GH expression or hGH expression from other transgenes.

REFERENCES

- Al-Shawi R Kinnaid, J., Burke, J., Bishop, J.O. 1990 Expression of a foreign gene in a line of transgenic mice is modulated by a chromosomal position effect. *Molecular and Cellular Biology* **10**: 1192-1198.

Altschul *et al.* (1994) *Nature Genetics* **6**:119-129.

- Ang H-L., Ivell, R., Walther, N., Nicholson, H., Ungefroren, H., Millar, M., Carter, D., Murphy, D. 1994 Over-expression of oxytocin in the testes of a transgenic mouse model. *Journal of Endocrinology* **140**: 53-62.

- Ang H-L, Ungefroren, H., De Bree, F., Foo, N.C., Carter, D., Burbach, J.P., Ivell, R., Murphy, D., 1991 Testicular oxytocin gene expression in seminiferous tubules of cattle and transgenic mice. *Endocrinology* **128**: 2110-2117.

- Banerjee SA, Roffler-Tarlov, S., Szabo, M., Frohman, L., Chikaraishi, D.M. 1994 DNA regulatory sequences of the rat tyrosine hydroxylase gene direct correct catecholaminergic cell-type specificity of a human growth hormone reporter in the CNS of transgenic mice causing a dwarf phenotype. *Molecular Brain Research* **24**: 89-106.

- Bartke, A., Steger, R.W., Hodges, S.L., Parkening, T.A., Collins, T.J., Yun, J.S., Wagner, T.E. 1988 Infertility in transgenic female mice and human growth hormone expression: evidence for luteal failure. *Journal of Experimental Zoology* **248**: 121-124.

- Bartke A., Naar, E.M., Johnson, L., May, M.R., Cecim, M., Yun, J.S., Wagner, T.E. 1992 Effects of expression of human or bovine growth hormone genes on sperm production and male reproductive performance in four lines of transgenic mice. *Journals of Reproduction and Fertility* **95**: 109-118.

- Bartke A., Cecim, M., Tang, K., Steger, R.W., Chandrashekar, V., Turyn, D. 1995 Neuroendocrine and reproductive consequences of overexpression of growth hormone in

transgenic mice. *Proceedings of the Society for Experimental Biology and Medicine* **206**: 345-359.

- 5 Belenky M., Castel, M., Young III, W.S., Gainer, H., Cohen, S. 1992 Ultrastructural immunolocalization of rat oxytocin-neurophysin in transgenic mice expressing the rat oxytocin gene. *Brain Research* **583**: 279-286.

- Bennett PA, Levy, A., Sophokleous, S., Robinson, I.C.A.F., Lightman, S.L. 1995 Hypothalamic GH receptor gene expression in the rat: effects of altered GH status.
10 *Journal Endocrinology* **147**: 225-234.

Birnboim HC, Doly, J. 1979 A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acid Research* **7**: 1513-1523.

- 15 Bonifer C Vidal, M., Grosveld, F., Sippel, A.E. 1990 Tissue-specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice. *The European Molecular Biology Organisation Journal* **9**: 2843-2848.

- Boss, O.; Samec, S.; Paoloni-Giacobino, A.; Rossier, C.; Dulloo, A.; Seydoux, J.;
20 Muzzin, P.; Giacobino, J.-P. : Uncoupling protein-3: a new member of the mitochondrial carrier family with tissue-specific expression. *FEBS Lett.* **408**: 39-42, 1997.

Bourne JA 1984 *Handbook of Immunoperoxidase Staining Methods*, DAKO corporation, Santa Barbara, USA.

25

Bray GA York, D.A. 1979 Hypothalamic and genetic obesity in experimental animals: an autonomic and endocrine hypothesis. *Physiological Reviews* **59**: 719-809.

- Brem G Wanke, R., Wolf, E., Buchmuller, T., Muller, M., Brenig, B., Hermanns, W.
30 1989 multiple consequences of human growth hormone expression in transgenic mice. *Molecular and Biological Medicine* **6**: 531-547.

Bucchini D Ripoche, M.-A., Stinnakre, M.-G., P.Desbois., Lores, P., Monthieux, E., Absil, J., Lepesant, J.-A., Pictet, R., Jami, J. 1986 Pancreatic expression of the human insulin gene in transgenic mice. *Proceedings of the National Academy of Science USA* **83**: 2511-2515.

5

Buijs RM 1987 Vasopressin localization and putative functions in the brain. *In* : Vasopressin principles and properties ; D.M. Gash and G.J. Boer (eds). 91-115.

10 Caffé AR van Leeuwen, F.W., Luiten, P.G.M. 1987 Vasopressin cells in the medial amygdala of the rat project to the lateral septum and ventral hippocampus. *The Journal of Comparative Neurology*. **261**: 237-252.

15 Cecim M Fadden, C., Kerr, J., Steger, R.W., Bartke, A. 1995 Infertility in transgenic mice overexpressing the bovine growth hormone gene: disruption of the neuroendocrine control of prolactin secretion during pregnancy. *Biology of Reproduction* **52**: 1187-1192.

20 Chandrashekar V Bartke, A., Wagner, T.E. 1991 Interactions of human growth hormone and prolactin on pituitary and Leydig cell function in adult transgenic mice expressing the human growth hormone gene. *Biology of Reproduction* **44**: 135-140.

Charlton HM Clark, R.G., Robinson, I.C.A.F., Porter Goff, A.E., Cox, B.S., Bugnon, C., Bloch, B.A. 1988 Growth hormone-deficient dwarfism in the rat: a new mutation. *Journal of Endocrinology* **119**: 51-58.

25 Chen, H., O. Charlat, L.A. Tartaglia, E.A. Woolf, X. Weng, S.J. Ellis, N.D. Lakey, J. Culpepper, K.J. Moore, R.E. Breitbart, G.M. Duyk, R.I. Tepper, and J.P. Morgenstern. 1996. Evidence that the diabetes gene encodes the leptin receptor: identification of a mutation in the leptin receptor gene in db/db mice. *Cell*. **84**:491-495.

30 Chomczynski P Sacchi, N. 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**: 156-9.

- Clark, R. G., Mortensen, D. L., Carlsson, L. M., Carlsson, B., Carmignac, D., and Robinson, I. C. A F. (1996). The obese growth hormone (GH)-deficient dwarf rat: body fat responses to patterned delivery of GH and insulin-like growth factor-I. *Endocrinology* 5 137, 1904-12.
- Coleman DL Eicher, E.M. 1990 Fat (fat) and tubby (tub): Two autosomal recessive mutations causing obesity syndromes in the mouse. *Journal of Hereditary* 81: 424-427.
- 10 Comuzzie, A.G. & Allison, D.B. (1998). The search for human obesity genes. *Science*, 280:1374-1377.
- Cool, D. R.; Normant, E.; Shen, F.; Chen, H.-C.; Pannell, L.; Zhang, Y.; Loh, Y. P. 1997: Carboxypeptidase E is a regulated secretory pathway sorting receptor: genetic obliteration 15 leads to endocrine disorders in Cpe(fat) mice. *Cell* 88: 73-83,
- De Lecea, L.; Kilduff, T. S.; Peyron, C.; Gao, X.-B.; Foye, P.E.; Danielson, P. E.; Fukuhara, C.; Battenberg, E. L. F.; Gautvik, V. T.; Bartlett, F. S., II; Frankel, W. N.; van den Pol, A. N.; Bloom, F. E.; Gautvik, K. M.; Sutcliffe, J. G. 1998 The hypocretins: 20 hypothalamus-specific peptides with neuroexcitatory activity. *Proc. Nat. Acad. Sci.* 95: 322-327.
- Dillon N Grosveld, F. 1993 Gene transcription - a practical approach. IRL Press at Oxford University Press. B.D. Hames and S.J. Higgins (eds): 153-187. 25
- Fairhall KM Carmignac, D.F., Robinson, I.C.A.F. 1992 Growth hormone (GH) binding protein and GH interactions in vivo in the guinea pig. *Endocrinology* 131: 1963-1969.
- Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J., and Cone, R. D. (1997). Role of 30 melanocortinerbic neurons in feeding and the agouti obesity syndrome. *Nature* 385, 165-8.

- Fitzsimmons MD Roberts, M.M., Robinson, A.G. 1994 Control of the posterior pituitary vasopressin content: Implications for the regulation of the vasopressin gene. *Endocrinology* **134**: 1874-1878.
- 5 Flavell DM., Wells, T., Wells, S.E., Carmignac, D.F., Thomas, G.B., Robinson, I.C.A.F. 1996 Dominant dwarfism in transgenic rats by targeting human growth hormone (GH) expression to hypothalamic GH-releasing factor neurons. *The European Molecular Biology Organisation Journal* **15**: 3871-3879.
- 10 Foo NC., Funkhouse, J.M., Carter, D., Murphy, D., 1994 A testis specific promoter in the rat vasopressin gene. *Journal of Biological Chemistry* **269**: 659-667.
- Fujiwara Y., Miwa, M., Takahashi, R., Hirabayashi, M., Suzuki, T., Ueda, M. 1997 Position-independent and high-level expression of human alpha-lactalbumin in the milk
15 of transgenic rats carrying a 210-kb YAC DNA construct. *Molecular Reproduction and development*. **47**: 157-163.
- Gainer H. & Wray, S. 1994 Cellular and molecular biology of oxytocin and vasopressin. In: *The Physiology of Reproduction*, E. Knobil and J.D. Neill (eds): 1099-112
20
- Good DJ., Porter, F.D., Mahon, K.A., Parlow, A.F., Weatphal, H., Kirsch, I.R. 1997 Hypogonadism and obesity in mice with a targeted deletion of the *Nhlh2* gene. *Nature Genetics* **15**: 397-401.
- 25 Gordon-Weeks, R 1987 PhD thesis. University of London.
- Graham, M.; Shutter, J. R.; Sarmiento, U.; Sarosi, I.; Stark, K. L. 1997.: Overexpression of *Agtr* leads to obesity in transgenic mice. *Nature Genet.* **17**: 273-274,
- 30 Grant SG., Jessee, J., Bloom, F.R., Hanahan, D. 1990 Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proceedings of the National Academy of Science* **87**: 4645-4649.

Grant FD., Reventos, J., Gordon, J.W., Kawabata, S., Miller, M., Majzoub, J.A. 1993a
Expression of the rat arginine vasopressin gene in transgenic mice. *Molecular
Endocrinology* 7: 659-667.

5

Grant FD., Reventos, J., Gordon, J.W., Kawabata, S., Miller, M., Majzoub, J.A. 1993b
Tissue-specific expression and osmotic regulation of a rat vasopressin gene in transgenic
mice. *Annals of the New York Academy of Sciences* 689: 530-533.

- 10 Gray RP & Yudkin, J.S. 1997 Cardiovascular disease in diabetes mellitus. In :
Textbook of diabetes, J. Pickup and G. Williams (eds): pp57.1-57.22.

- Grosveld F., van Assendelft, G.B., Greaves, D.R., Kollias, G. 1987 Position-
independent, high-level expression of human β -globin gene in transgenic mice. *Cell* 51:
15 975-985.

Habener JF., Cwikel, B.J., Hermann, H., Hammer, R.E., Palmiter, R.D., Brinster, R.L. 1989
Metallothionein-vasopressin fusion gene expression in transgenic mice. *The Journal of
Biological chemistry* 264: 18844-18852.

20

Hammer RE., Pursel, V.G., Rexroad, C.E., Wall, R.J., Bolt, D.J., Ebert, K.M., Palmiter,
R.D., Brinster, R.L. 1985 Production of transgenic rabbits, sheep and pigs by
microinjection. *Nature* 315: 680-683.

- 25 Hanahan D. 1983 Studies on transformation of *Escherichia coli* with plasmids. *Journal
of Molecular Biology* 166: 557-569.

Ho M-Y., Carter, D.A., Ang, H-L., Murphy, D. 1995 Bovine oxytocin transgenes in
mice. *Journal of Biological Chemistry*. 270: 27199-27205.

30

Hogan B., Constantini, F., Lacy, E. 1986 Manipulating the mouse embryo. Cold Spring
Harbor Laboratory Press, Cold Spring Harbor, N.Y.

- Hollingshead PG., Martin, L., Pitts, S.L., Stewart, T.A. 1989 A dominant phenocopy of hypopituitarism in transgenic mice resulting from central nervous system synthesis of human growth hormone. *Endocrinology* **125**: 1556-1564.
- 5
Horn AM., Robinson, I.C.A.F., Fink, G. 1985 Oxytocin and vasopressin in rat hypophysial portal blood: experimental studies in normal and Brattleboro rats. *Journal of Endocrinology* **104**: 211-224.
- 10 Huber MC., Bosch, F.X., Sippel, A.E., Bonifer, C. 1994 Chromosomal position effect in chicken lysozyme gene transgenic mice are correlated with suppression of DNase I hypersensitive site formation. *Nucleic Acid Research* **22**: 4195-4201.
- Huszar, D., Lynch, C. A., Fairchild, H. V., Dunmore, J. H., Fang, Q., Berkemeier, L. R.,
15 Gu, W., Kesterson, R. A., Boston, B. A., Cone, R. D., Smith, F. J., Campfield, L. A., Burn, P., and Lee, F. (1997). Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131-41.
- Iida M., Murakami, T., Ishida, K., Mizuno, A., Kuwajima, M., Shima, K. 1996
20 Substitution at codon 269 (glutamine-proline) of the leptin receptor (OF-R) cDNA is the only mutation found in the Zucker Fatty (fa/fa) rat. *Biochemical and Biophysical Research Communications* **224**: 597-604.
- Ikeda, A. Chang, K., Matsuyama, M., Nishihara, M. & Takahashi, M. 1995 Disorders in
25 carbohydrate metabolism in two lines of transgenic rats expressing different levels of human growth hormone. *Proceedings of US Endocrine Society Washington* P2-241.
- Ikeda, A., Matsumoto, Y., Chang, K.T., Nakano, T., Matsuyama, S., Yamanouchi, K.,
Ohta, A., Nishihara, M., Tojo, H., Sasaki, F. & Takahashi, M (1997). Different female
30 reproductive phenotypes determined by human growth hormone (hGH) levels in hGH-transgenic rats. *Biol Reprod.* **56**: 847-851.

- Ikeda, A., Matsuyama, S. Nishihara, M., Tojo, H. & Takahashi, M. 1994. Changes in endogenous growth hormone secretion and onset of puberty in transgenic rats expressing human growth hormone gene. *Endocr J.* **41**: 523-529.
- 5 Ivell R & Richter, D. 1984 Structure and comparison of the oxytocin and vasopressin genes from rat. *Proceedings of the National Academy of Science, USA.* **81**: 2006-2010.
- Jamal, Z., Martin, A., Gomez, M. A., Hales, P., Chang, E., Russell, J. C., and Brindley, D. N. (1997). Phosphatidate phosphohydrolases in liver, heart and adipose tissue of the
- 10 JCR:LA corpulent rat and the lean genotypes: implications for glycerolipid synthesis and signal transduction. *Int J Obes Relat Metab Disord* **16**, 789-99.
- Jones BK., Monks, B.R., Liebhaber, S.A., Cooke, N.E. 1995 The human growth hormone gene is regulated by multicomponent locus control region. *Molecular and*
- 15 *Cellular Biology* **15**: 7010-7021.
- Kahle, E. B., Butz, K. G., Chua, S. C., Kershaw, E. E., Leibel, R. L., Fenger, T. W., Hansen, C. T., and Michaelis, O. E. (1997). The rat corpulent (cp) mutation maps to the same interval on (Pgm1-Glut1) rat chromosome 5 as the fatty (fa) mutation. *Obes Res* **5**,
- 20 142-5.
- Kiyama H & Emson, P.C. 1990 Evidence for the co-expression of oxytocin and vasopressin messenger ribonucleic acids in magnocellular neurosecretory cells: simultaneous demonstration of two neurohypophysin messenger ribonucleic acids by
- 25 hybridization histochemistry. *Journal of Neuroendocrinology* **2**: 257-259.
- Klebig, M. L., Wilkinson, J. E., Geisler, J. G., and Woychik, R. P. (1995). Ectopic expression of the agouti gene in transgenic mice causes obesity, features of type II diabetes, and yellow fur. *Proc Natl Acad Sci U S A* **92**, 4728-32.
- 30 Kleyn, P. W., Fan, W., Kovats, S. G., Lee, J. J., Pulido, J. C., Wu, Y., Berkemeier, L. R., Misumi, D. J., Holmgren, L., Charlat, O., Woolf, E. A., Tayber, O., Brody, T., Shu, P.,

- Hawkins, F., Kennedy, B., Baldini, L., Ebeling, C., Alperin, G. D., Deeds, J., Lakey, N. D., Culpepper, J., Chen, H., Glucksmann, K. M., Moore, K. J., et al (1996). Identification and characterization of the mouse obesity gene *tubby*: a member of a novel gene family. *Cell* **85**, 281-90.
- 5 Kristensen, P.; Judge, M. E.; Thim, L.; Ribel, U.; Christjansen, K. N.; Wulff, B. S.; Clausen, J. T.; Jensen, P. B.; Madsen, O. D.; Vrang, N.; Larsen, P. J.; Hastrup, S. (1998). Hypothalamic CART is a new anorectic peptide regulated by leptin. *Nature* **393**: 72-76
- 10 Lacy E., Roberts, S., Evans, E.P., Burtenshaw, M.D., Costantini, F.D. 1983 A foreign β -globin gene in transgenic mice: Integration at abnormal chromosomal positions and expression in inappropriate tissues. *Cell* **34**: 343-358.
- Land H, Graz,M, Ruppert,S, Schmale,H, Rehbein,M, Richter,D, Schutz,G, 1983
- 15 Deduced amino acid sequence from the bovine oxytocin-neurophysin precursor cDNA. *Nature* **302**: 342-344.
- Lee & Costlow 1987. *Methods in Enzymology* **152**:633-648.
- 20 Lee, G., Li, C., Montez, J., Halaas, J., Darvishzadeh, J., and Friedman, J. M. (1997). Leptin receptor mutations in 129 db3J/db3J mice and NIH fap/fap rats. *Mamm Genome* **8**, 445-7.
- Lightman SL & Young III, W.S. 1987 Vasopressin, oxytocin, dynorphin, enkephalin
- 25 and corticotrophin-releasing factor mRNA stimulation in the rat. *Journal of Physiology* **394**: 23-39.
- Mansour *et al.*, (1989) *Nature* **336**, 348-352.
- 30 Mathe, D. (1995). Dyslipidemia and diabetes: animal models. *Diabetes Metab* **21**, 106-11.

- McGrane MM, de Vente, J., Yun, J., Bloom, J., Park, E., Wynshaw-Boris, A., Wagner, T., Rottman, F.M., Hanson, R.W. 1988 Tissue-specific expression and dietary regulation of a chimeric phosphoenolpyruvate carboxykinase/ bovine growth hormone gene in transgenic mice. *Journal of Biological Chemistry* **263**: 11443-11451.
- 5
McNamee, C. J., Kappagoda, C. T., Kunjara, R., and Russell, J. C. (1994). Defective endothelium-dependent relaxation in the JCR:LA-corpulent rat. *Circ Res* **74**, 1126-32.
- 10 Michaelis, O. Velasquez, M. T., Abraham, A. A., Servetnick, D. A., Scholfield, D. J., and Hansen, C. T. (1995). Development and characteristics of a new strain of obese hyperinsulinemic and hyperlipidemic Dahl salt-sensitive rat. The Dahl salt-sensitive/NIH-corpulent rat. *Am J Hypertension*.
- 15 Miller MW., Duhl, D.M., Vrieling, H., Cordes, S.P., Ollmann, M.M., Winkes, B.M., Barsh, G.S. 1993 Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutations. *Genes and Development* **7**: 454-467.
- 20 Millet, L.; Vidal, H.; Andreelli, F.; Larrouy, D.; Riou, J.-P.; Ricquier, D.; Laville, M.; Langin, D. 1997 Increased uncoupling protein-2 and -3 mRNA expression during fasting in obese and lean humans. *J. Clin. Invest.* **100**: 2665-2670.
- 25 Mohr E., Schmitz, E., Richter, D. 1988 A single rat genomic DNA fragment encodes both the oxytocin and vasopressin genes separated by 11 kilobases and orientated in opposite transcriptional directions. *Biochimie* **70**: 649-654.
- Moon, B. C., and Friedman, J. M. (1997). The molecular basis of the obese mutation in ob2J mice. *Genomics* **42**, 152-6.

Morello D., Moore, G., Salmon, A. M., Yaniv, M., Babinet, C. 1986 Studies on the expression of an H-2K/human growth hormone fusion gene in giant transgenic mice. The European Molecular Biology Organisation Journal 5: 1877-1883.

- 5 Mullins JJ., Peters, J., Ganten, D., 1990 Fulminant hypertension in transgenic rats harbouring the mouse Ren-2 gene. Nature 344: 541-544.

Murphy D. & Carter, D., 1990 Vasopressin gene expression in the Rodent hypothalamus: Transcriptional and posttranscriptional responses to physiological stimulation. Molecular
10 Endocrinology 1054: 1051-1059.

Murphy D & Ho, M-Y. 1995 Oxytocin transgenic mice. Advances in Experimental Medicine and Biology 395: 67-78.

- 15 Naggert, J. K.; Fricker, L. D.; Varlamov, O.; Nishina, P. M.; Rouille, Y.; Steiner, D. F.; Carroll, R. J.; Paigen, B. J.; Leiter, E. H. 1995. Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity. Nature Genet. 10: 135-142

- 20 Ninomiya, T., Hirabayashi, M., Sagara, J. & Yuki, A. 1994 Functions of milk protein gene 5' flanking regions on human growth hormone gene. Mol Reprod Dev 37: 276-283.

North, M. A., Naggert, J. K., Yan, Y., Noben, T. K., and Nishina, P. M. (1997). Molecular
25 characterization of TUB, TULP1, and TULP2, members of the novel tubby gene family and their possible relation to ocular diseases. Proc Natl Acad Sci U S A 94, 3128-33.

Ohki-Hamazaki, H., Watase, K., Yamamoto, K., Ogura, H., Yamano, M., Yamada, K., Maeno, H., Imaki, J., Kikuyama, S., Wada, E., and Wada, K. (1997). Mice lacking bombesin receptor subtype-3 develop metabolic defects and obesity. Nature 390, 165-169.

- Ollmann, M. M.; Wilson, B. D.; Yang, Y.-K.; Kerns, J. A.; Chen, Y.; Gantz, I.; Barsh, G. S. 1995 Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. *Science* **278**: 135-138, 1997.
- 5 Ornitz DM., Palmiter, R.D., Hammer, R.E., Brinster, R.L., Swift, G.H., MacDonald, R.J. 1985 Specific expression of an elastase-human growth hormone fusion gene in pancreatic acinar cells of transgenic mice. *Nature* **313**: 600-602.
- Palmiter RD., Brinster, R.L., Hammer, R.E., Trumbauer, M.E., Rosenfeld, M.G.,
- 10 Birnberg, N.C., Evans, R.M. 1982 Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature* **300**: 611-615.
- Palmiter RD., Norstedt, G., Gelinas, R.E., Hammer, R.E., Brinster, R.L. 1983 Metallothionein-human GH fusion genes stimulate growth of mice. *Science* **222**: 809-
- 15 814.
- Pursel VG., Pinkert, C. A., Miller, K. F., Bolt, D. J., Campbell, R. G., Palmiter, R. D., Brinster, R. L., Hammer, R. E. 1989 Genetic engineering of livestock. *Science* **244**: 1281-1287.
- 20 Qu, D.; Ludwig, D. S.; Gammeltoft, S.; Piper, M.; Pelleymounter, M. A.; Cullen, M. J.; Mathes, W. F.; Przypek, J.; Kanarek, R.; Maratos-Flier, E. (1996). A role for melanin-concentrating hormone in the central regulation of feeding behaviour. *Nature* **380**: 243-247.
- 25 Quaife CJ., Mathews, L.S., Pinkert, C.A., Hammer, R.E., Brinster, R.L., Palmiter, R.D. 1989 Histopathology associated with elevated levels of growth hormone and insulin-like growth factor I in transgenic mice. *Endocrinology* **124**: 40-48.
- 30 Reaven, GR. 1988 Role of insulin resistance in human disease. *Diabetes* **37**: 1595-1607.

- Richard, D., Rivest, R., Naimi, N., Timofeeva, E., and Rivest, S. (1996). Expression of corticotropin-releasing factor and its receptors in the brain of lean and obese Zucker rats. *Endocrinology* **137**, 4786-95.
- 5 Roskam WG., Rougeon, F. 1979 Molecular cloning and nucleotide sequence of the human growth hormone structural gene. *Nucleic Acid Research* **10**: 305-320.
- Russell, J. C., Amy, R. M., Graham, S., Wenzel, L. M., and Dolphin, P. J. (1993). Effect of castration on hyperlipidemic, insulin resistant JCR:LA-corpulent rats. *Atherosclerosis*
 10 **100**, 113-22.
- Russo AF., Crenshaw, E.B, Lira, S.A, Simmons, D.M, Swanson, L.W, Rosenfeld, M.G 1988 Neuronal expression of chimeric genes in transgenic mice. *Neuron* **1**: 311-320.
- 15 Sakurai, T.; Amemiya, A.; Ishii, M.; Matsuzaki, I.; Chemelli, R. M.; Tanaka, H.; Williams, S. C.; Richardson, J. A.; Kozlowski, G. P.; Wilson, S.; Arch, J. R. S.; Buckingham, R. E.; Haynes, A. C.; Carr, S. A.; Annan, R. S.; McNulty, D. E.; Liu, W.-S.; Terrett, J. A.; Elshourbagy, N. A.; Bergsma, D. J.; Yanagisawa, M. 1998. Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors
 20 that regulate feeding behavior. *Cell* **92**: 573-585
- Sambrook J., Fritsch E.F., Maniatis, T. 1989 *Molecular Cloning. A Laboratory Manual*: second edition.
- 25 Sanger F., Nicklen, S., Coulson, A.R. 1977 DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Society of Science. USA* **74**: 5463-5467.
- Sausville E, Carney D, Battey J 1985 The human vasopressin gene is linked to the oxytocin gene and is selectively expressed in a cultured lung cancer cell line. *J Biol Chem*
 30 **260**: 10236-10241

- Schmitz E., Mohr, E., Richter, D., 1991 Rat vasopressin and oxytocin genes are linked by a long interspersed repeated DNA element (LINE): Sequence and Transcriptional analysis of LINE. *DNA and Cell Biology* **10**: 81-91.
- 5 Seeburg PH., Shine, J., Martial, J.A., Baxter, J.D., Goodman, H.M. 1977 Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. *Nature* **270**: 486-494.
- Shanahan CM., Rigby, N.W., Murray, J.D., Marshall, J.T., Townrow, C.A., Nancarrow, C.D., Ward, K.A. 1989 Regulation of expression of a sheep metallothionein 1a-Sheep growth hormone fusion gene in transgenic mice. *Molecular and Cellular Biology* **9**: 5473-5479.
- 10
- Shea BT., Hammer, R.E., Brinster, R.L. 1987 Growth allometry of the organs in giant transgenic mice. *Endocrinology* **121**: 1924-1930.
- 15
- Shillabeer, G., Hornford, J., Forden, J. M., Wong, N. C., Russell, J. C., and Lau, D. C. (1992). Fatty acid synthase and adipsin mRNA levels in obese and lean JCR:LA-cp rats: effect of diet. *J Lipid Res* **33**, 31-9.
- 20
- Short MK., Clouthier, D.E., Schaefer, I.M., Hammer, R.E., Magnuson, M.A., Beale, E.G. 1992 Tissue-specific, developmental, hormonal and dietary regulation of rat phosphoenolpyruvate carboxykinase-human growth hormone fusion genes in transgenic mice. *Molecular and Cellular Biology* **12**: 1007-1020.
- 25
- Southern EM. 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**: 503-509.
- Stewart TA., Clift, S., Pitts-Meek, S., Martin, L., Terrell, T.G., Liggitt, D., Oakley, H. 1992 An evaluation of the functions of the 22-kilodalton (kDa), the 20 kDa, and the N-terminal polypeptide forms of human growth hormone using transgenic mice. *Endocrinology* **130**: 405-414.
- 30

- Swanson L.W., Simmons, D.M., Arriza, J., Hammer, R., Brinster, R., Rosenfeld, M.G., Evans, R.M. 1985 Novel developmental specificity in the nervous system of transgenic animals expressing growth hormone fusion genes. *Nature* **317**: 363-366.
- 5 Szabo M., Butz, M.R., Banerjee, S.A., Chikaraishi, D.M., Frohman, L.A. 1995 Autofeedback suppression of growth hormone (GH) secretion in transgenic mice expressing a human GH reporter targeted by tyrosine hydroxylase 5'-flanking sequences to the hypothalamus. *Endocrinology* **136**: 4044-4048.
- 10 Takaya K., Ogawa, Y., Isse, N., Okazaki, T., Satoh, Masuzaki, H., Mori, K., Tamura, N., Hosoda, K., Nakao, K. 1996 Molecular cloning of rat leptin isoform complementary DNAs- Identification of a missense mutation in Zucker fatty (fa/fa) rats. *Biochemical and Biophysical Research Communications* **225**: 75-83.
- 15 Takeuchi, T., Suzuki, H., Sakurai, S., Nogami, H., Okuma, S., and Ishikawa, H. (1990). Molecular mechanism of growth hormone (GH) deficiency in the spontaneous dwarf rat - detection of abnormal splicing of GH messenger ribonucleic acid by the polymerase chain reaction. *Endocrinology* **126**, 31-38.
- 20 Takiguchi S, Takata Y, Takahashi N, Kataoka K, Hirashima T, Kawano K, Miyasaka K, Funakoshi A, Kono A. 1998 A disrupted cholecystokinin A receptor gene induces diabetes in obese rats synergistically with ODB1 gene. *Am J Physiol* **274**:E265-E270
- 25 Tartaglia L.A., Dembski, M., Weng, X., Deng, N., Culpepper, J., Devos, R., Richards, G.J., Campfield, L.A., Clark, F.T., Deeds, J., Muir, C., Sanker, S., Moriarty, A., Moore, E.A., Monroe, C.A., Tepper, R.I. 1995 Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**: 1263-1271.
- 30 Taylor, W.R., 1998 Dynamic databank searching with templates and multiple alignment. *J. Mol. Biol.* **280**: 375-406.

Urban JH., Miller, M.A., Drake, C.T., Dorsa, D.M. 1990 Detection of vasopressin mRNA in cells of the medial amygdala but not the locus coeruleus by in situ hybridization. *Journal of Chemical Neuroanatomy* 3: 277-283.

5

Vandesande P. & Dierickx, D. 1975 Identification of the vasopressin producing and of the oxytocin producing neurons in the hypothalamic magnocellular neurosecretory system of the rat. *Cell and Tissue Research* 164: 153-162.

- 10 Van Tol HHM., Bolwerk, E.L.M., Lui, B., Burbach, J.P. 1988 Oxytocin and vasopressin gene expression in the hypothalamo-neurophyseal system of the rat during the estrous cycle, pregnancy and lactation. *Endocrinology* 122: 945-951.

- 15 Vidal-Puig, A., Solanes, G., Grujic, D., Flier, J. S & Lowell, B. B. 1997 UCP3: an uncoupling protein homologue expressed preferentially and abundantly in skeletal muscle and brown adipose tissue. *Biochem. Biophys. Res. Commun.* 235: 79-82.

- 20 Wahl GM., Lewis, K.A., Ruiz, J.C., Rothenberg, B., Zhao, J., Evans, G.A. 1987 Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer. *Proceeding of the National Academy of Science. USA* 84: 2160-2164.

Waller S., Fairhall, K. M., Xu, J., Robinson, I. C. A. F., Murphy, D. 1996 Neurohypophyseal and fluid homeostasis in transgenic rats expressing a tagged rat vasopressin prepropeptide in hypothalamic neurons. *Endocrinology* 137: 5068-5077.

25

Whitaker, R. C., Wright, J. A., Pepe, M. S., Seidel, K. D., and Dietz, W. H. (1997). Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med* 337, 869-73.

- 30 White, RB, Eisen, JA, Kasten, TL & Fernald, RD 1998 Second gene for gonadatropin-releasing hormone in humans *PNAS* 95, 305-309.

- Wu-Peng XS, Chua SC Jr, Okada N, Liu SM, Nicolson M, Leibel RL 1997 Phenotype of the obese Koletsky (f) rat due to Tyr763Stop mutation in the extracellular domain of the leptin receptor (Lepr): evidence for deficient plasma-to-CSF transport of leptin in both the Zucker and Koletsky obese rat. *Diabetes* 46:513-518
- 5
- Young III WS., Reynolds,K, Shepard,E.A, Gainer,H, Castel,M, 1990 Cell-specific expression of the rat oxytocin gene in transgenic mice. *Journal of Neuroendocrinology* 2: 917-925.
- 10
- Young III WS 1992 Expression of oxytocin and vasopressin genes. *Journal of Neuroendocrinology* 4: 527-540.
- Yun JS., Li, Y.S., Wight, D.C., Portanova, R., Selden, R.F. 1989 The human growth hormone transgene : expression in hemizygous and homozygous mice. *Proceedings of the Society for Experimental Biology and Medicine* 194: 308-313.
- 15
- Zeng Q., Carter, D.D., Murphy, D. 1994b Cell specific expression of a vasopressin transgene in rats. *Journal of Neuroendocrinology* 6: 469-477.
- 20
- Zeng Q., Foo, N-C., Funkhouser, J. M., Carter, D. A., Murphy, D. 1994a Expression of a rat vasopressin transgene in rat testes. *Journal of Reproduction and Fertility* 102: 471-481.
- 25
- Zhang Y., Proenca, R., Maffei, M., Barone, M., Leopold, L., Friedman, J.M. 1994 Positional cloning of the mouse obese gene and its human homologue. *Nature* 372: 425-427.

Claims

1. 5'OT-EST polypeptide having a sequence selected from the group comprising the sequences set forth in any one of SEQ. ID. Nos. 2, 4 or 6, and sequences substantially
5 homologous to any one of the polypeptides set forth in SEQ. ID. Nos. 2, 4 or 6.

2. A polypeptide according to claim 1 comprising an amino acid sequence encoded by at least one exon selected from the group consisting of exons w, x, y and z as set forth in SEQ. ID. No. 16, or equivalents thereof as set forth in any one of SEQ. ID. Nos. 3 or 5.
10

3. A polypeptide according to claim 2, which comprises an amino acid sequence encoded by at least part of exon w as set forth in SEQ. ID. No. 16, or equivalents thereof as set forth in any one of SEQ. ID. Nos. 3 or 5.

15 4. A mutant of a 5'OT-EST polypeptide according to any preceding claim which is capable, *in vivo*, of modulating the obesity of an animal expressing it.

5. A mutant according to claim 4, wherein the animal is a transgenic animal expressing the mutant as a result of transformation with a transgene.
20

6. A mutant according to claim 4 or claim 5, which comprises the sequence PRPRSFSAFSSQDS, or a sequence substantially homologous thereto.

7. A mutant according to any one of claims 4 to 6 which comprises the sequence
25 MLRALNRLAARPGGQPPTLLLLPVRGPRPRSFSAFSSQDS, or a sequence substantially homologous thereto.

8. A nucleic acid encoding a 5'OT-EST polypeptide or mutant 5'OT-EST polypeptide according to any preceding claim.
30

9. A nucleic acid according to claim 8, having a sequence selected from the group consisting of any one of SEQ. ID. Nos. 1, 3, 5, 7, 16 or 17; sequences which are

hybridisable under stringent conditions with an oligonucleotide comprising 20 contiguous bases from any one of SEQ. ID. Nos. 1, 3, 5, 7, 16 or 17; sequences substantially homologous to any one of SEQ. ID. Nos. 1, 3, 5, 7, 16 or 17; and sequences complementary thereto.

5

10. A nucleic acid according to claim 9, comprising the sequence ATGTTGCGGGCTT TGAACCGCCTGGCCGCGCGGCGGGGCCAGCCCCAACCCCTGCTCCTTCTGCCCCTGC GCGGCCCCACGGCCCCGCTCATTCTCGGCTCCTTTTTCCTCGCAGGATAGC, or an equivalent sequence which encodes the same polypeptide having regard to the degeneracy of the nucleic acid code, or a sequence substantially homologous thereto.

10

11. A nucleic acid vector comprising a nucleic acid sequence according to any one of claims 8 to 11.

15

12. A vector according to claim 11 which is a cosmid vector.

13. A vector according to claim 11 or claim 12 further comprising the sequences of the oxytocin (OT) gene, the vasopressin (AVP) gene and/or the human growth hormone (hGH) gene.

20

14. A vector according to claim 12 having the structure of cVO14 as set forth in Figure 4 (SEQ. ID. No. 17).

15. A cell transformed with a vector according to any one of claims 11 to 13.

25

16. A method for producing a 5'OT-EST polypeptide or a mutant 5'OT-EST polypeptide according to any one of claims 1 to 7, comprising transforming a cell with a vector according to any one of claims 11 to 13 and culturing the cell to produce the polypeptide.

30

17. A transgenic non-human animal expressing, as a result of transgene expression, a 5'OT-EST polypeptide or mutant 5'OT-EST polypeptide according to any one of claim 1 to 7.

5 18. A transgenic animal according to claim 17, which has been transformed with a vector according to any one of claims 12 to 14.

19. A transgenic animal according to claim 17 or claim 18, comprising more than one copy of the transgene.

10

20. A transgenic animal according to any one of claims 17 to 19, which is a mammal.

21. A transgenic animal according to claim 20 which is a rat.

15 22. A transgenic rat comprising at least four concatameric copies of a transgene having the structure of cVO14 as set forth in Figure 4 (SEQ. ID. No. 17).

20 23. A non-human mammal possessing the following obese phenotype: (i) a very late onset of obesity, (ii) a highly selective visceral distribution of fat developing on a normal rodent diet, without hyperphagia, (iii) an effect greatly preponderant in males, (iv) a predisposition to excessive dietary-fat induced obesity at an early age, before the phenotype becomes apparent on a normal diet, and (v) a dominant pattern of inheritance; the non-human mammal being obtainable by transformation with a vector according to any one of claims 11 to 14.

25

24. Use of an animal according to any one of claims 17 to 23 as a model for human late onset obesity, human dietary-fat associated juvenile obesity, human female post-menopausal obesity and/or human male infertility.

30 25. A method for identifying a compound or compounds capable of modulating obesity and/or infertility in a mammal, comprising the steps of:

a) exposing an animal according to any one of claims 17 to 24 to the compound or compounds to be tested;

b) determining the effect of the compound on the obesity and/or infertility phenotype; and

5 c) selecting the compound or compounds which are capable of modulating the obesity and/or infertility phenotype in the desired manner.

26. A method for producing a compound or compounds capable of modulating obesity and/or infertility in a mammal, comprising the steps of:

10 a) exposing an animal according to any one of claims 17 to 24 to the compound or compounds to be tested;

b) determining the effect of the compound on the obesity and/or infertility phenotype;

15 c) selecting the compound or compounds which are capable of modulating the obesity and/or infertility phenotype in the desired manner; and

d) producing the compound or compounds by conventional isolation or synthesis techniques.

27. A method for identifying a candidate compound capable of influencing lipid transport, comprising the steps of:

a) contacting 5'OT-EST polypeptide with a candidate compound or compounds and determining which candidate compound or compounds is capable of interacting with 5'OT-EST;

25 b) optionally, testing candidate compounds which interact with 5'OT-EST in a method according to claim 25.

28. A diagnostic reagent for the detection of mutations, polymorphisms or other changes in 5'OT-EST which may predispose an individual to obesity.

30 29. Use of a tissue derived from a transgenic animal according to any one of claims 17 to 24 in a screen to identify a genetic cause of obesity, comprising the steps of:

a) isolating one or more gene products from tissue derived from a transgenic animal according to any one of claims 17 to 24; and

b) determining whether the expression of a gene product is correlated with obesity.

Abstract

The invention describes a previously unknown gene, termed *5'OT-EST*, which is responsible for inducing an obesity and/or infertility phenotype in transgenic animals, and
5 transgenic animals comprising mutants of *5'OT-EST* which are useful for assaying compounds for the treatment of obesity and/or infertility.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: MEDICAL RESEARCH COUNCIL
- (B) STREET: 20 PARK CRESCENT
- (C) CITY: LONDON
- (E) COUNTRY: UK
- (F) POSTAL CODE (ZIP): W1N 4AL

(ii) TITLE OF INVENTION: GENE

(iii) NUMBER OF SEQUENCES: 16

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 924 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 5..604

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

TGTC ATG TTG CGG GCT TTG AAC CGC CTG GCC GCG CGG CCC GGG GGC CAG	49
Met Leu Arg Ala Leu Asn Arg Leu Ala Ala Arg Pro Gly Gly Gln	
1 5 10 15	
CCC CCA ACC CTG CTC CTT CTG CCC GTG CGC GGC CGC AAG ACC CGC CAC	97
Pro Pro Thr Leu Leu Leu Leu Pro Val Arg Gly Arg Lys Thr Arg His	
20 25 30	
GAT CCG CCT GCC AAG TCC AAG GTC GGG CGC GTG AAA ATG CCT CCT GCA	145
Asp Pro Pro Ala Lys Ser Lys Val Gly Arg Val Lys Met Pro Pro Ala	
35 40 45	
GTG GAC CCT GCG GAA TTG TTC GTG TTG ACC GAG CGC TAC CGA CAG TAC	193
Val Asp Pro Ala Glu Leu Phe Val Leu Thr Glu Arg Tyr Arg Gln Tyr	
50 55 60	

CGG GAG ACG GTG CGC GCT CTC AGG CGA GAG TTC ACA TTG GAG GTG CGA Arg Glu Thr Val Arg Ala Leu Arg Arg Glu Phe Thr Leu Glu Val Arg 65 70 75	241
GGG AAA TTG CAC GAG GCC CGA GCC GGG GTT CTG GCT GAG CGC AAG GCG Gly Lys Leu His Glu Ala Arg Ala Gly Val Leu Ala Glu Arg Lys Ala 80 85 90 95	289
CAA GAG GCC ATC AGA GAG CAC CAG GAG CTG ATG GCC TGG AAC CGG GAG Gln Glu Ala Ile Arg Glu His Gln Glu Leu Met Ala Trp Asn Arg Glu 100 105 110	337
GAG AAC CGG AGA CTG CAG GAA CTA CGG ATA GCT AGG TTG CAG CTC GAA Glu Asn Arg Arg Leu Gln Glu Leu Arg Ile Ala Arg Leu Gln Leu Glu 115 120 125	385
GCA CAG GCC CAG GAG CTG CGG CAG GCT GAG GTC CAG GCC CAG AGG GCC Ala Gln Ala Gln Glu Leu Arg Gln Ala Glu Val Gln Ala Gln Arg Ala 130 135 140	433
CAG GAG GAG CAG GCT TGG GTG CAA CTG AAA GAA CAA GAA GTT CTC AAA Gln Glu Glu Gln Ala Trp Val Gln Leu Lys Glu Gln Glu Val Leu Lys 145 150 155	481
CTG CAG GAG GAG GCC AAA AAC TTC ATC ACT CGG GAG AAC CTG GAG GCA Leu Gln Glu Glu Ala Lys Asn Phe Ile Thr Arg Glu Asn Leu Glu Ala 160 165 170 175	529
CGG ATA GAA GAG GCC TTG GAC TCT CCG AAG AGT TAT AAC TGG GCG GTC Arg Ile Glu Glu Ala Leu Asp Ser Pro Lys Ser Tyr Asn Trp Ala Val 180 185 190	577
ACC AAA GAA GGG CAG GTG GTC AGG AAC TGAGAACAGA GGCCTCTCAG Thr Lys Glu Gly Gln Val Val Arg Asn 195 200	624
GCCCAAATAA GGACAGTGCT TGCCTAGGGA CTGGATATTG GGGTAGAAAT TGGTGCATCC	684
CAGGAGGGTG GCACAGCCTT GTCCAGAGCA GCCCCCATTC ATTCTAGATT TGGCACCAGG	744
TATAGTACCT GTTCTGACAC CACATACAAA CTCCGGACAG CATTAAACTC TGGGAAGTTC	804
CTATCACACA GAAGATCAGA CTGGACTGTC CCCTCTAGAA GCCAAGAGCT GTCTCCTGAG	864
TTTCTTGAA TAGTGTGAGC CCAATGTTTC CTGCTTTTAT AAATAAACTA TTGGAAAGCA	924

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 200 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Leu Arg Ala Leu Asn Arg Leu Ala Ala Arg Pro Gly Gly Gln Pro
1 5 10 15

3

Pro Thr Leu Leu Leu Leu Pro Val Arg Gly Arg Lys Thr Arg His Asp
 20 25 30

Pro Pro Ala Lys Ser Lys Val Gly Arg Val Lys Met Pro Pro Ala Val
 35 40 45

Asp Pro Ala Glu Leu Phe Val Leu Thr Glu Arg Tyr Arg Gln Tyr Arg
 50 55 60

Glu Thr Val Arg Ala Leu Arg Arg Glu Phe Thr Leu Glu Val Arg Gly
 65 70 75 80

Lys Leu His Glu Ala Arg Ala Gly Val Leu Ala Glu Arg Lys Ala Gln
 85 90 95

Glu Ala Ile Arg Glu His Gln Glu Leu Met Ala Trp Asn Arg Glu Glu
 100 105 110

Asn Arg Arg Leu Gln Glu Leu Arg Ile Ala Arg Leu Gln Leu Glu Ala
 115 120 125

Gln Ala Gln Glu Leu Arg Gln Ala Glu Val Gln Ala Gln Arg Ala Gln
 130 135 140

Glu Glu Gln Ala Trp Val Gln Leu Lys Glu Gln Glu Val Leu Lys Leu
 145 150 155 160

Gln Glu Glu Ala Lys Asn Phe Ile Thr Arg Glu Asn Leu Glu Ala Arg
 165 170 175

Ile Glu Glu Ala Leu Asp Ser Pro Lys Ser Tyr Asn Trp Ala Val Thr
 180 185 190

Lys Glu Gly Gln Val Val Arg Asn
 195 200

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 998 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..615

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATG CTA CGC GCG CTG AGC CGC CTG GGC GCG GGG ACC CCG TGC AGG CCC
 Met Leu Arg Ala Leu Ser Arg Leu Gly Ala Gly Thr Pro Cys Arg Pro

205	210	215	
CGG GCC CCT CTG GTG CTG CCA GCG CGC GGC CGC AAG ACC CGC CAC GAC Arg Ala Pro Leu Val Leu Pro Ala Arg Gly Arg Lys Thr Arg His Asp 220 225 230			96
CCG CTG GCC AAA TCC AAG ATC GAG CGA GTG AAC ATG CCG CCC GCG GTG Pro Leu Ala Lys Ser Lys Ile Glu Arg Val Asn Met Pro Pro Ala Val 235 240 245			144
GAC CCT GCG GAG TTC TTC GTG CTG ATG GAG CGT TAC CAG CAC TAC CGC Asp Pro Ala Glu Phe Phe Val Leu Met Glu Arg Tyr Gln His Tyr Arg 250 255 260			192
CAG ACC GTG CGC GCC CTC AGG ATG GAG TTC GTG TCC GAG GTG CAG AGG Gln Thr Val Arg Ala Leu Arg Met Glu Phe Val Ser Glu Val Gln Arg 265 270 275 280			240
AAG GTG CAC GAG GCC CGA GCC GGG GTT CTG GCG GAG CGC AAG GCC CTG Lys Val His Glu Ala Arg Ala Gly Val Leu Ala Glu Arg Lys Ala Leu 285 290 295			288
AAG GAC GCC GCC GAG CAC CGC GAG CTG ATG GCC TGG AAC CAG GCG GAG Lys Asp Ala Ala Glu His Arg Glu Leu Met Ala Trp Asn Gln Ala Glu 300 305 310			336
AAC CGG CGG CTG CAC GAG CTG CGG ATA GCG AGG CTG CGG CAG GAG GAG Asn Arg Arg Leu His Glu Leu Arg Ile Ala Arg Leu Arg Gln Glu Glu 315 320 325			384
CGG GAG CAG GAG CAG CGG CAG GCG TTG GAG CAG GCC CGC AAG GCC GAA Arg Glu Gln Glu Gln Arg Gln Ala Leu Glu Gln Ala Arg Lys Ala Glu 330 335 340			432
GAG GTG CAG GCC TGG GCG CAG CGC AAG GAG CGG GAA GTG CTG CAG CTG Glu Val Gln Ala Trp Ala Gln Arg Lys Glu Arg Glu Val Leu Gln Leu 345 350 355 360			480
CAG GAA GAG GTG AAA AAC TTC ATC ACC CGA GAG AAC CTG GAG GCA CGG Gln Glu Glu Val Lys Asn Phe Ile Thr Arg Glu Asn Leu Glu Ala Arg 365 370 375			528
GTG GAA GCA GCA TTG GAC TCC CGG AAG AAC TAC AAC TGG GCC ATC ACC Val Glu Ala Ala Leu Asp Ser Arg Lys Asn Tyr Asn Trp Ala Ile Thr 380 385 390			576
AGA GAG GGG CTG GTG GTC AGG CCA CAA CGC AGG GAC TCC TAGGGGCCCA Arg Glu Gly Leu Val Val Arg Pro Gln Arg Arg Asp Ser 395 400 405			625
GTAAGGACAG TGCCCGCCAG GGACCATGTA TGTATCATGG CGGAAGAGTT GGCCCTGACC			685
TGGAATAAAG CAGTTGGTGT TGCTTATGAG GAAGGTTTCAG CCTTATCCAG CACAGCCTTC			745
ACGTTTTGCC CTCTGCTGTC ACCACTTGGT CAGAACTTC CAAACGCAGT GCCCTGTTCT			805
GCCGGTGTGT AAAGCCTCAG CGCACCAGGA GACCCTAGAG TGGTTTCCAT CTCACAGAGA			865
ATCAGACAGG CCACAGCCCC CTCAGGCAGC CAGGTCATCT GAGTATCATT AAGAGTAGTG			925
ATGGGAAGAT TACAGTCTGA GGGCCAAACG TGCCTGCTTC CTGTTTTTGT AAATAAAGTT			985

TTGTTGGAAC ACA

998

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 205 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

Met Leu Arg Ala Leu Ser Arg Leu Gly Ala Gly Thr Pro Cys Arg Pro
 1              5              10              15

Arg Ala Pro Leu Val Leu Pro Ala Arg Gly Arg Lys Thr Arg His Asp
          20              25              30

Pro Leu Ala Lys Ser Lys Ile Glu Arg Val Asn Met Pro Pro Ala Val
          35              40              45

Asp Pro Ala Glu Phe Phe Val Leu Met Glu Arg Tyr Gln His Tyr Arg
          50              55              60

Gln Thr Val Arg Ala Leu Arg Met Glu Phe Val Ser Glu Val Gln Arg
          65              70              75              80

Lys Val His Glu Ala Arg Ala Gly Val Leu Ala Glu Arg Lys Ala Leu
          85              90              95

Lys Asp Ala Ala Glu His Arg Glu Leu Met Ala Trp Asn Gln Ala Glu
          100             105             110

Asn Arg Arg Leu His Glu Leu Arg Ile Ala Arg Leu Arg Gln Glu Glu
          115             120             125

Arg Glu Gln Glu Gln Arg Gln Ala Leu Glu Gln Ala Arg Lys Ala Glu
          130             135             140

Glu Val Gln Ala Trp Ala Gln Arg Lys Glu Arg Glu Val Leu Gln Leu
          145             150             155             160

Gln Glu Glu Val Lys Asn Phe Ile Thr Arg Glu Asn Leu Glu Ala Arg
          165             170             175

Val Glu Ala Ala Leu Asp Ser Arg Lys Asn Tyr Asn Trp Ala Ile Thr
          180             185             190

Arg Glu Gly Leu Val Val Arg Pro Gln Arg Arg Asp Ser
          195             200             205

```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 943 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

6

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 5..604

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

TGTC ATG TTG CGC GCT CTG AAC CGC CTG GCG CAG CGG CCG GGA GAC CGG	49
Met Leu Arg Ala Leu Asn Arg Leu Ala Gln Arg Pro Gly Asp Arg	
210 215 220	
CCC CCG ACC CCG CTG CTC CTG CCC GTG CGC GGC CGC AAG ACC CGC CAT	97
Pro Pro Thr Pro Leu Leu Leu Pro Val Arg Gly Arg Lys Thr Arg His	
225 230 235	
GAC CCG CCT GCC AAA TCC AAG GTC GGA CGG GTG CAG ACG CCT CCC GCC	145
Asp Pro Pro Ala Lys Ser Lys Val Gly Arg Val Gln Thr Pro Pro Ala	
240 245 250	
GTG GAC CCT GCG GAA TTC TTC GTG TTG ACC GAG CGC TAC GGA CAG TAC	193
Val Asp Pro Ala Glu Phe Phe Val Leu Thr Glu Arg Tyr Gly Gln Tyr	
255 260 265	
CGG GAG ACC GTG CGC GCT CTC AGG CTA GAG TTC ACG TTG GAT GTG CGA	241
Arg Glu Thr Val Arg Ala Leu Arg Leu Glu Phe Thr Leu Asp Val Arg	
270 275 280	
AGG AAA TTG CAC GAG GCC CGA GCC GGG GTT CTG GCC GAG CGC AAG GCG	289
Arg Lys Leu His Glu Ala Arg Ala Gly Val Leu Ala Glu Arg Lys Ala	
285 290 295 300	
CAG CAG GCC ATC ACG GAG CAC CGG GAG CTG ATG GCC TGG AAC CGG GAC	337
Gln Gln Ala Ile Thr Glu His Arg Glu Leu Met Ala Trp Asn Arg Asp	
305 310 315	
GAG AAC CGG CGA ATG CAG GAG CTA CGG ATA GCG AGG TTG CAG CTG GAA	385
Glu Asn Arg Arg Met Gln Glu Leu Arg Ile Ala Arg Leu Gln Leu Glu	
320 325 330	
GCA CAG GCC CAG GAG GTG CAG AAG GCT GAG GCC CAG CGC CAG AGG GCT	433
Ala Gln Ala Gln Glu Val Gln Lys Ala Glu Ala Gln Arg Gln Arg Ala	
335 340 345	
CAG GAG GAG CAG GCT TGG GTG CAA CTG AAA GAG CAA GAA GTG CTC AAG	481
Gln Glu Glu Gln Ala Trp Val Gln Leu Lys Glu Gln Glu Val Leu Lys	
350 355 360	
CTG CAG GAG GAG GCA AAA AAC TTC ATC ACT CGG GAG AAC CTG GAG GCA	529
Leu Gln Glu Glu Ala Lys Asn Phe Ile Thr Arg Glu Asn Leu Glu Ala	
365 370 375 380	
CGG ATA GAA GAA GCG TTG GAC TCT CCG AAG AGT TAC AAC TGG GCC GTC	577
Arg Ile Glu Glu Ala Leu Asp Ser Pro Lys Ser Tyr Asn Trp Ala Val	
385 390 395	

ACC AAA GAA GGG CAG GTG GTC AGG AAC TGAGCACAGA GACTTCTGGG 624
 Thr Lys Glu Gly Gln Val Val Arg Asn
 400 405

GGCCCAAATA AGCACAGTGC TTGCCTAGGG TCTGTGTACT GGGATAGGAA TTGGTACATC 684
 CCAGGAGGAT GGCTCAGCCG TTTCCAGAGC AACCTCAGTC ACTCCAGGCT CGGCACTCAC 744
 CACCTGACTG GGAAC TCCCA GATGTCCCTG TTCTGGCACC ACAGTCAAAC TGAGGGCAGC 804
 ATTAAACTCT GGGAAGTTCC TATCGCACAG AGGATCGGAC TGGACTGTGT CCCTCTAGAA 864
 GCCAAGCTTG TCTTGTAAGT CTCTGGAGT CCTGTGAGCC AAATGTTTCC TGCTTTTATA 924
 AATAAAGTAT TGGAGCCCA 943

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 200 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Leu Arg Ala Leu Asn Arg Leu Ala Gln Arg Pro Gly Asp Arg Pro
 1 5 10 15

Pro Thr Pro Leu Leu Pro Val Arg Gly Arg Lys Thr Arg His Asp
 20 25 30

Pro Pro Ala Lys Ser Lys Val Gly Arg Val Gln Thr Pro Pro Ala Val
 35 40 45

Asp Pro Ala Glu Phe Phe Val Leu Thr Glu Arg Tyr Gly Gln Tyr Arg
 50 55 60

Glu Thr Val Arg Ala Leu Arg Leu Glu Phe Thr Leu Asp Val Arg Arg
 65 70 75 80

Lys Leu His Glu Ala Arg Ala Gly Val Leu Ala Glu Arg Lys Ala Gln
 85 90 95

Gln Ala Ile Thr Glu His Arg Glu Leu Met Ala Trp Asn Arg Asp Glu
 100 105 110

Asn Arg Arg Met Gln Glu Leu Arg Ile Ala Arg Leu Gln Leu Glu Ala
 115 120 125

Gln Ala Gln Glu Val Gln Lys Ala Glu Ala Gln Arg Gln Arg Ala Gln
 130 135 140

Glu Glu Gln Ala Trp Val Gln Leu Lys Glu Gln Glu Val Leu Lys Leu
 145 150 155 160

Gln Glu Glu Ala Lys Asn Phe Ile Thr Arg Glu Asn Leu Glu Ala Arg
 165 170 175

8

Ile Glu Glu Ala Leu Asp Ser Pro Lys Ser Tyr Asn Trp Ala Val Thr
 180 185 190

Lys Glu Gly Gln Val Val Arg Asn
 195 200

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2852 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vii) IMMEDIATE SOURCE:

(B) CLONE: Rat 5'OT-EST-xdel

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION:1026..1270

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION:1799..2235

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:1030..1152

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TGACCTCTGT GGATCTGATA TACATGTAAG TGACAGACCA TCCGAGCTAT ATAGTGAGAC	60
CTGTGCAAGG AAGGATGGAG TGCACGTTCC CTGATGTTCA GAGCAACCCT GTGTCACTCC	120
AGGTAGGTGA GATGAGAGGA AGAGGGTGGC CTTGGCCTGG GCCTCCTACG GGCCTGGAAG	180
TTGGGAGAAG GATGTAAGCA GACTCTGTTC TCTTCTGAGA AATATCAGGT ATTGCAGTCA	240
GCCCAGGCTC CTCAGACCCT CCTAAGTGCA GATTCTCTGC AGAATCTGGT GTTGACAACA	300
CTAATGAGTA GGATGAGACT TCAGTTCCCT AGCCCTCACC GTCAGCTTCT GATTACCAAC	360
AACTCTCCCA GAGGAGAGCC ATCTACCTTT GGGACAGATG CTCTCTGCCC TGCCTGCCT	420
CCTGTTTCTC TTCATTGTAG AGGAAGATAG TACTTTAAAA GCTTCATAAA TGGTCTCAAG	480
GTGGGAAGAC CCCGGCTCAG GTGAAAGAGG ACAAGCGTCA CCTCACACAG GCCACCCAGT	540
AGAAAACAAG TGATCACTGA TACTGAGAAC TCTGGCAATT GCAGAGCTGC CCAAGACCAC	600
AACAGGGCAG TGCAATGCAA GGAAAAGGTT TGTGCTCGA TTGCAAACCT AAAGTTTAAA	660

GTGCATCAGG AGAACGCTTA CTCAAAGAGG AAGTGTAAGC CTAACCTAAG TAGCTAGAAG	720
CTCAGAATTT CTTGCATCAG CCCTGGAAGG GTACACAGGC CACCGGTGGG CCAGAGAACC	780
ACACGCTTTG GGGCGGTGTC CAAGCTTGTG AACAAAGTAGG CAAGAGCGCC TGGTGTGTGA	840
GCTGTCATTG GCGGGCAATA CAGCCCAGCG AACTGTGGTC TCCAAGGTGC CCCTCGACCC	900
TCCCACTCTA CCCGAGACTC CAGGGACGCG ATGGGCCAGA CAGCAAGAGC TCCGCCTACG	960
GGGGCGGGGA CAGGAGATTC CCGTGATGCT CCTCGACCAC TTCCGGACAG GGCGCAGGCG	1020
CTAGCTGTC ATG TTG CGG GCT TTG AAC CGC CTG GCC GCG CGG CCC GGG	1068
Met Leu Arg Ala Leu Asn Arg Leu Ala Ala Arg Pro Gly	
205 210	
GGC CAG CCC CCA ACC CTG CTC CTT CTG CCC GTG CGC GGC CCA CGG CCC	1116
Gly Gln Pro Pro Thr Leu Leu Leu Leu Pro Val Arg Gly Pro Arg Pro	
215 220 225	
CGC TCA TTC TCG GCT CCT TTT TCC TCG CAG GAT AGC TAGGTTGCAG	1162
Arg Ser Phe Ser Ala Pro Phe Ser Ser Gln Asp Ser	
230 235 240	
CTCGAAGCAC AGGCCCCAGGA GCTGCGGCAG GCTGAGGTCC AGGCCCAGAG GGCCCAGGAG	1222
GAGCAGGCTT GGGTGCAACT GAAAGAACAA GAAGTTCTCA AACTGCAGGT GGGCCGAGGT	1282
CGTGAGGAAT GTGGGTATTG GAGATTCCGG TGAGGGAGGC TCTGGGGAGA GCAGCACAGG	1342
GTGTCAAGTG ACCAGTCTTC AGGAGGCTTC TCTCTCTGCT CTGCACACAC AGAGTGCCTC	1402
CCAGACAATG GTCAATGAAA GGTTACAGGC TAGTATTGCC GTGTGAAACT TGAAGGTCAG	1462
GGAAACCATA AATGAGAATG GAGCTGTTTT TATTGTGTAA GGGAGAGTGA CAAGGTTGAG	1522
AGAGTCCACC ACCCCGCACC TCCCCCGCC CCCAATCAGG TTGTCACGAT TCGATTTCGT	1582
CTTGGGTTGT GGCTGAGAGA TCTGATGGGT AATTGTCCGA GGAAGAGGGA TATAATGGTT	1642
GAGGTCACTT AGTACAGTTG TGCTGGCCTA TTGGTGGGAC ACTCAAAGGG GCCCTGGGCT	1702
CTTTTGACAC CCTTCTTAAG GTGGGCTAGA GACAGTAAGT TATGCAGGCA GCCAGCTCTG	1762
AGAGATCCCA CGTAGCTAAC CTTTCTCTTC CCGTAGGAGG AGGCCAAAAA CTTCATCACT	1822
CGGGAGAACC TGGAGGCACG GATAGAAGAG GCCTTGGA CTCCGAAGAG TTATAACTGG	1882
GCGGTCACCA AAGAAGGGCA GGTGGTCAGG AACTGAGAAC AGAGGCCTCT CAGGCCCAAA	1942
TAAGGACAGT GCTTGCCTAG GGACTGGATA TTGGGGTAGA AATTGGTGCA TCCCAGGAGG	2002
GTGGCACAGC CTTGTCCAGA GCAGCCCCCA TTCATTCTAG ATTTGGCACC AGGTATAGTA	2062
CCTGTTCTGA CACCACATAC AAACCTCCGGA CAGCATTAATA CTCTGGGAAG TTCCTATCAC	2122
ACAGAAGATC AGACTGGACT GTCCCCTCTA GAAGCCAAGA GCTGTCTCCT GAGTTTCTTG	2182
GAATAGTGTG AGCCCAATGT TTCCTGCTTT TATAAATAAA CTATTGGAAA GCAAAGCCTT	2242
TTGTTATGTG GCTTGCTTTT TCTTGTTGTA GAATAAGTTT ATTTGTCCCA GTTATTTGGG	2302

TCTTAAGGTT ATTAGCCAAA AGCCAGTTCA CCTAACTGAG CCAGGAGTTA GTTATCTGCT 2362
 TTGCTCAATC CTGGGCTTTG CTGGGTAGGG TCAGGTGTGT CCAAGGTCCA GAAAGCAAAA 2422
 AGGGTGCCCC GTTCTCCTG GGAAGGCTTC CCCGTCAGTG ATTTCTGTAA CCGGACCCTG 2482
 CCCTGACACA GCGTCATTGG ACTACCCAGC AGACAGTAGA CTCCACTCTA AACCCGCTTC 2542
 TTGCGGTCAG TTGCTGTCCT TCAGTGTGTG TAAGCAGTGG CCAGACAGCA CCCTTGGGTG 2602
 TCATTTCAAG ACTCTCTCAC CTTGGTCTGC TTTACGTTTG GTTTGATTTG GTTTGTTCTG 2662
 GTTTTTGAGA CGAGGCCTTT CACTGGAACC TGGCACTCAG TATTTAGACT GCCCAGCCAG 2722
 CTAGCCTCAG AGAATGCATC TGCATATGCT TGCCTGGCGC TGAATTCGG TGCACATGGC 2782
 TTTGATGTGT ACCGGGGATC AGACACAGAT GTTTCATGAG TGCAGTGCAT GCCTGTTAGT 2842
 GGTAGAGCTC 2852

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Leu Arg Ala Leu Asn Arg Leu Ala Ala Arg Pro Gly Gly Gln Pro
 1 5 10 15
 Pro Thr Leu Leu Leu Leu Pro Val Arg Gly Pro Arg Pro Arg Ser Phe
 20 25 30
 Ser Ala Pro Phe Ser Ser Gln Asp Ser
 35 40

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

- (A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TTCACACCAC TCTGTCGAAC

20

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AGGAGGAAGA CAGGTGAAAG

20

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TCATGTTGCG GGCTTTGAAC

20

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TCTTTCAGTT GCACCCAAGC

20

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GTGATAGGAA CTTCCCAGAG

20

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

GCCTCGTGCA ATTTCCCTCG CACCTCCAAT GTGAACTCTC GC

42

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TCCTGCGAGG AAAAAGGAGC CGAGAATGAG CGGGGCCGTG GG

42

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3264 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vii) IMMEDIATE SOURCE:
 (B) CLONE: Rat 5'OT-EST

(ix) FEATURE:
 (A) NAME/KEY: exon w
 (B) LOCATION:1026..1241

(ix) FEATURE:
 (A) NAME/KEY: exon x
 (B) LOCATION:1332..1478

(ix) FEATURE:
 (A) NAME/KEY: exon y
 (B) LOCATION:1559..1682

(ix) FEATURE:
 (A) NAME/KEY: exon z
 (B) LOCATION:2211..2647

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TGACCTCTGT GGATCTGATA TACATGTAAG TGACAGACCA TCCGAGCTAT ATAGTGAGAC	60
CTGTGCAAGG AAGGATGGAG TGCACGTTCC CTGATGTTCA GAGCAACCCT GTGTCACTCC	120
AGGTAGGTGA GATGAGAGGA AGAGGGTGGC CTTGGCCTGG GCCTCCTACG GGCCTGGAAG	180
TTGGGAGAAG GATGTAAGCA GACTCTGTTC TCTTCTGAGA AATATCAGGT ATTGCAGTCA	240
GCCCAGGCTC CTCAGACCCT CCTAAGTGCA GATTCTCTGC AGAATCTGGT GTTGACAACA	300
CTAATGAGTA GGATGAGACT TCAGTTCCCT AGCCCTCACC GTCAGCTTCT GATTACCAAC	360

AACTCTCCCA	GAGGAGAGCC	ATCTACCTTT	GGGACAGATG	CTCTCTGCCC	TGCACTGCCT	420
CCTGTTTCTC	TTCATTGTAG	AGGAAGATAG	TACTTTAAAA	GCTTCATAAA	TGGTCTCAAG	480
GTGGGAAGAC	CCCGGCTCAG	GTGAAAGAGG	ACAAGCGTCA	CCTCACACAG	GCCACCCAGT	540
AGAAAACAAG	TGATCACTGA	TACTGAGAAC	TCTGGCAATT	GCAGAGCTGC	CCAAGACCAC	600
AACAGGGCAG	TGCAATGCAA	GGAAAAGGTT	TGTTGCTCGA	TTGCAAACCT	AAAGTTTAAA	660
GTGCATCAGG	AGAACGCTTA	CTCAAAGAGG	AAGTGTAAGC	CTAACTTAAG	TAGCTAGAAG	720
CTCAGAATTT	CTTGCACTAG	CCCTGGAAGG	GTACACAGGC	CACCGGTGGG	CCAGAGAACC	780
ACACGCTTTG	GGGCGGTGTC	CAAGCTTGTG	AACAAGTAGG	CAAGAGCGCC	TGGTGTGTGA	840
GCTGTCATTG	GCGGGCAATA	CAGCCCAGCG	AACTGTGGTC	TCCAAGGTGC	CCCTCGACCC	900
TCCCACTCTA	CCCAGAGACTC	CAGGGACGCG	ATGGGCCAGA	CAGCAAGAGC	TCCGCCTACG	960
GGGGCGGGGA	CAGGAGATTG	CCGTGATGCT	CCTCGACCAC	TTCCGGACAG	GGCGCAGGCG	1020
CTAGCTGTCA	TGTTGCGGGC	TTTGAACCGC	CTGGCCGCGC	GGCCCGGGGG	CCAGCCCCCA	1080
ACCCTGCTCC	TTCTGCCCCG	GCGCGGCCGC	AAGACCCGCC	ACGATCCGCC	TGCCAAGTCC	1140
AAGGTCGGGC	GCGTGAAAAT	GCCTCCTGCA	GTGGACCCTG	CGGAATTGTT	CGTGTTGACC	1200
GAGCGCTACC	GACAGTACCG	GGAGACGGTG	CGCGCTCTCA	GGTGTGTGTA	AAGGGCAGGC	1260
GGCCTTCGGC	GCCCCCTGGG	AAGTGCTGGG	GCTGGAGGAT	GGGTGCTCAC	TTGAAGCCCG	1320
TCCTCACCCA	GGCGAGAGTT	CACATTGGAG	GTGCGAGGGA	AATTGCACGA	GGCCCGAGCC	1380
GGGGTTCTGG	CTGAGCGCAA	GGCGCAAGAG	GCCATCAGAG	AGCACCAGGA	GCTGATGGCC	1440
TGGAACCGGG	AGGAGAACCG	GAGACTGCAG	GAACTACGGT	GCGAGAGGCG	CGGGGCTGGG	1500
TGGGCTGGGC	TAGGCTCACC	CACGGCCCCG	CTCATTCTCG	GCTCCTTTTT	CCTCGCAGGA	1560
TAGCTAGGTT	GCAGCTCGAA	GCACAGGCCC	AGGAGCTGCG	GCAGGCTGAG	GTCCAGGCCC	1620
AGAGGGCCCA	GGAGGAGCAG	GCTTGGGTGC	AACTGAAAGA	ACAAGAAGTT	CTCAAAGTGC	1680
AGGTGGGCCG	AGGTCGTGAG	GAATGTGGGT	ATTGGAGATT	CCGGTGAGGG	AGGCTCTGGG	1740
GAGAGCAGCA	CAGGGTGTCA	AGTGACCAGT	CTTCAGGAGG	CTTCTCTCTC	TGCTCTGCAC	1800
ACACAGAGTG	CCTCCAGAC	AATGGTCAAT	GAAAGGTTAC	AGGCTAGTAT	TGCCGTGTGA	1860
AACTTGAAGG	TCAGGGAAAC	CATAAATGAG	AATGGAGCTG	TTTTTATTGT	GTAAGGGAGA	1920
GTGACAAGGT	TGAGAGAGTC	CACCACCCCG	CACCTCCCCC	CGCCCCCAAT	CAGGTTGTCA	1980
CGATTGATT	CGTTCTTGGG	TTGTGGCTGA	GAGATCTGAT	GGGTAATTGT	CCGAGGAAGA	2040
GGGATATAAT	GGTTGAGGTC	ACCTAGTACA	GTTGTGCTGG	CCTATTGGTG	GGACACTCAA	2100
AGGGGCCCTG	GGCTCTTTTG	ACACCCTTCT	TAAGGTGGGC	TAGAGACAGT	AAGTTATGCA	2160
GGCAGCCAGC	TCTGAGAGAT	CCCACGTAGC	TAACCTTTCT	CTTCCCGTAG	GAGGAGGCCA	2220

AAAACTTCAT CACTCGGGAG AACCTGGAGG CACGGATAGA AGAGGCCTTG GACTCTCCGA	2280
AGAGTTATAA CTGGGCGGTC ACCAAAGAAG GGCAGGTGGT CAGGAACTGA GAACAGAGGC	2340
CTCTCAGGCC CAAATAAGGA CAGTGCTTGC CTAGGGACTG GATATTGGGG TAGAAATTGG	2400
TGCATCCCAG GAGGGTGGCA CAGCCTTGTC CAGAGCAGCC CCCATTCATT CTAGATTTGG	2460
CACCAGGTAT AGTACCTGTT CTGACACCAC ATACAAACTC CGGACAGCAT TAACTCTGG	2520
GAAGTTCCCTA TCACACAGAA GATCAGACTG GACTGTCCCC TCTAGAAGCC AAGAGCTGTC	2580
TCCTGAGTTT CTTGGAATAG TGTGAGCCCA ATGTTTCCTG CTTTATATAA TAACTATTG	2640
GAAAGCAAAG CCTTTTGTTA TGTGGCTTGC TTTTCTTGT TGTAGAATAA GTTTATTTGT	2700
CCCAGTTATT TGGGTCTTAA GGTTATTAGC CAAAAGCCAG TTCACCTAAC TGAGCCAGGA	2760
GTTAGTTATC TGCTTTGCTC AATCCTGGGC TTTGCTGGGT AGGGTCAGGT GTGTCCAAGG	2820
TCCAGAAAGC AAAAAGGGTG CCCCCTTTCT CCTGGGAAGG CTTCCCCGTC AGTGATTTCT	2880
GTAACCGGAC CCTGCCCTGA CACAGCGTCA TTGGACTACC CAGCAGACAG TAGACTCCAC	2940
TCTAAACCCG CTTCTTGCGG TCAGTTGCTG TCCTTCAGTG TGTGTAAGCA GTGGCCAGAC	3000
AGCACCCCTG GGTGTCATTT CAAGACTCTC TCACCTTGGT CTGCTTTACG TTTGGTTTGA	3060
TTTGGTTTGT TCTGGTTTTT GAGACGAGGC CTTTCACTGG AACCTGGCAC TCAGTATTTA	3120
GACTGCCCAG CCAGCTAGCC TCAGAGAATG CATCTGCGTA TGCTTGCCTG GCGCTGGAAT	3180
TCGGTGCACA TGGCTTTGAT GTGTACCGGG GATCAGACAC AGATGTTTCA TGAGTGCAGT	3240
GCATGCCTGT TAGTGGTAGA GCTC	3264

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 44576 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: circular
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Cosmid DNA"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GCGGCCGCAT AATACGACTC ACTATAGGGA TCTGGTGGAG GACCTATGGC CCGCGAGCTA	60
---	----

GAGAAGTGGT TCTCAACCTT CCTAGTGCTG AGACCCTTTA ACACAGTTCC TCGTGTTGTG	120
GGGAAACCCC CTCCTGCAAC CATAAAATAA TTTTGTGTTAC TACTTCATAA CAAGTGTTGC	180
TACTCTATTG CTATGAATTG TAAAATAAAT GTGTCTTCCA ATGGTCTTAG ATGACTCCCG	240
TGAAAGGGTC ATTCTACCCC TAAGAGGTCA TGATCTACAG GTTGAGAACC ACTGATCTCC	300
AGTAACCTTC ACTTGAGTCC ATATCCTCCA TGAAGGTATG GAAGTCAATA AAACTGAGCT	360
TCAAGCCTCA TCAAAATGGG TCCATCCCCT GGTACAGTGT GAGTGGAAGA ATACCCACCA	420
TACGGTCACT GGAAGGAGGA TGTCTGAAGG GTCTTAGATT GTGTCAAGGG GTCCTGGGTG	480
TCAGGATCTG ACGAAGCAGG CTCGTCATGT TTCATGAAGA CTACAGGTAT GTGATAAAAC	540
TGCAAGCTGG AAAAGTACCC ACTGAGCCCG TGTGGCTCTG CTGGGATTTG GAGGCATGAG	600
GAGCAGAGGG TCTGGAGGAC AGCAGTCCCA GAAATAATCT ATGACTAAGA AGGCTGAACT	660
GGGGTGA CTC TCTGGTGGA AGAGTTGCCT TTAAAGAAGG AAGACATACC AGGCATAGCA	720
ACAACTGCCT TTAGTACTAG CACTCTGAAG GCAGAGGAAG TCCGATTTCT CTGAGTTCCA	780
AGCCAGCTTG GTTTACACAG CAAGTTCTAG GCCAACTAGG GTTACATAGT GGA CTCTCCT	840
CAAACGGGGT TGAGAAAGGA CTCAGCAGTT AGCTCAGTTA ACTCCAGTTC TAGGAAATAT	900
GATCCCTTAG TCTGACCTCT TGGCATGTAA GTGGTGCACA TACATATATG CACACAAAAT	960
ACATCAATCT GCAAAGGGGG AGGGAGGAAG GGCTGGAGTC TGAAGAAATA GTTCAGTGGT	1020
TAAGAGAATT CACTGCTCTT CCCAATAGCC AAATTCAGCT CCTAGCATCC ATGTCAGATG	1080
GCCCACGAAC ACCTGTAATT CTAGCCCCTA AACTCAGTGC CCCTTCACAA GACGGGGACA	1140
CACGTACACA TATACCTAAA AAATTAGGTG GTTTTTTTTT ATTTATAAGG TCAAATGCAG	1200
AATATCAAAT GGGTTAGACA GCAGCTCCAA GCTGGCCTCT TCCTCCCAGG GCTCTTCTTG	1260
ACTCTTGGCA CCCTCTTTGG GTCCAGAACC CAGACATTAG CCATGACTCA GCTGATAAAA	1320
TGCAACCCAT GGCTCATTA TTAGGAAGTC TGTAATTAGC CTGTCTGGTA GCCTCCAGAG	1380
AGAACCCCTT TCACCTGTCT TCCTCCTCTC ACCCAGGGGA AGAGCTCAGT TTTGCCCCTG	1440
AGACAGAAGA AGGGAACGAG ACCATGAGCA ACGGGAAATG AGATGCTGGC GCACACACAC	1500
TTTATGTGTG TGAAGTCTCA GAGAGGTCAC CAATAATGAG GCAATGGAAA TGAGCTGAGC	1560
TGCCTGAACC TCCAAGTTTC CTCCAAGAAA ACCCCACAGG GGAGATGGGG CATGGCCCAG	1620
GCCAGCTGCC CCAGCCTCTG CTGGCAGAAA GTGAGCCCGC TGCCATTTTA ATTTTGTATA	1680
CAGGGTCTCA CTCTACAGCT CTGGGGGCCT AAAACTCACT ATGTAGACTT CAAACTCAAC	1740
CAAACCAACA ACAAAAACAA ACAAACCCC TGCACTGACT GGAGAGATGG CTCGGTTGAG	1800
AACAATGGCT GCTAGGAGTC AAACCCAGGT CCTGTGGAAG AGCATGCTGG TAACTGCTGG	1860

GTCATCGCTG	GGTCACTCTC	TTCACACACA	CACACACACA	CACACACACA	CACGGCAATG	1920
AACTCTTCAG	TGTCTTGATT	TACGGTTTCT	TCCGATAAAT	CCTCAGGAGG	GCAGTCAAGT	1980
GGCTCATTTG	GCAAATGCTT	GCCTGAGACC	TGAGTTTGGT	TCCCAGAACC	CATGGAGGCA	2040
GAAGGAAAGG	GCTCCACAAA	GCTCTCTTCT	GAACTCCATA	TGTGCACACA	CACCCACTTC	2100
GCACACATTC	ATAATAGTGA	TGAATGAAAA	TGAAGACAGA	TAAAAAAAAC	CAATTTCTGTG	2160
AAACTGTTAG	CACGTTTCAGT	CAATGGCTTT	GGGGGTAAAC	TGTTTCAGAG	CCATGGTACT	2220
CAGTCACTAG	GCTCATACTG	GTCAGACGCT	GAGGTCAGCA	ATGGAGAGCT	GCTACACCTA	2280
AAGGTAGCAG	AGGTCATTTG	GCTCTGACTC	AGAATATTCC	AGCTCTCCAC	ATTCACAGAA	2340
GTTCTACTTG	GTCGTAGAAA	AAAGCTGAGC	CTTTTTTTTT	TTTGGAAC	TTATTTTTTT	2400
AAAGATATAT	TTATTTTATG	TATATGAGTG	CACTGTAGCT	GTCTTCAGAC	ACACCAGAAG	2460
GGGGCATCGG	ATCCCATTAC	AGATGGCTGT	GAGCCAACAT	GTGGTCGCTG	GGGATTGAAC	2520
TTAGGACCTC	TGGAAGAGCA	GTCAGTGCTC	TTAACCGCTG	AGCCATCTCT	CCAGCCCTGG	2580
AACTTTATTT	TGAACATGCA	ACCCACCTA	CCACTATGGG	TTCAGTCACC	AGCGCCTTAG	2640
GAATAAAATT	GGAGAAAATA	AGCTTTATGG	TTAGTCAGCT	GTCAGCTGTG	GGGTTGGGGA	2700
CAGAAGAATG	GTTATGTTTT	GTTTTCCCAT	CAAGGCCTCA	CTCTGTGACC	TGGTTGGTCT	2760
GCCACTTGCT	CTGTAGATCA	GGTTTCAATT	ACAGAGATCC	ACCTGCTCCG	TGTCGCTATG	2820
CTGGGATAAG	AACTAAGTCA	CCCTGCCTAC	CTTATTACTT	TGTATTCTTG	GGCATGGAAC	2880
TCAATTCCTT	GTCAGCGAGA	GAATAACTTC	CTCGATCGGA	GTGTTTTTAT	GTGAATTGGG	2940
CCAAAAAGAC	TGCGATGCTC	TGAGACCTAT	TTGTGAAGCC	AAGAGTAGTG	GGTAGCACAA	3000
GTCAGAAATC	CAAGGACTTG	GTAAGCTGAG	ACAGTAGGAT	GTGTGCGCTC	ATGCACACAC	3060
ACACACACAG	ACACACACAC	AGACACATAC	ATGCATGGAC	GCACAGAGGC	ACCCACGCAC	3120
ATGTGCCTGG	ATGAGGCTTC	AGTTCTTCAT	AAAGCTGCCT	TTGAGTTTGT	GCCCTCCCAC	3180
TCTTCCTGAG	GACTIONGAGTC	CTCACACCTT	GGGCTGATAG	TGCACCACTA	CCTTTTTTTAG	3240
TGACCTCCTC	TTTGCAGTCA	CAGGCTGAAG	GTACAGGGAG	GACTIONGAGTC	GCCGTCTGCC	3300
TCTGTTTAAC	ATGAACCTGC	AAGGCAGTGG	GCAGCCTCAC	CCCTAGCGAT	GGCACTIONGAGT	3360
GATGCCAGGA	ACGCTGTCCT	CATGTGCCCT	TGGCTGTTGG	GGCACAGTGT	GCCTCTGCAG	3420
GGCCAGCCTG	ACCGTGTGTG	CCAGCCAGAA	TGCACAATTT	CTGCCCCGACC	TTGGAAGCTT	3480
TTTGTCTTTC	CTTGTGAGTT	TCTTGTCAAC	CAGCAGTGTT	TCTTGCCTCT	TTGCTTGACG	3540
CCTCTATGGG	AAGATGGACA	AGACTTTTTT	TTTTCTACAT	CCCCTGCAAA	CAGGTTTGTC	3600
ATACCTCTCA	GGGGCAGGGG	TCTTGTCCCT	GTCAAGCGCA	GCAGGCCACC	AGACCCAGAA	3660
CTATGAAATC	TACCCAACTT	GTCTCTGTAC	AAAGTTAAAC	AACAAAAAGA	AACTTGTTTT	3720

TGTTTTTGT	TTTTTTTTTG	TTTTGTTTTG	TTTTGTTTTT	TGAGACAGGG	TTTCTCTATG	3780
TAGTCCTGGC	TATTCTGGAA	CTTGTTCTAT	AGACCAGGCT	ATCCTGGAAC	TCAAAGAACG	3840
GCCTGACTCT	GTCTCCCGGG	TGCTGGTCAC	TCTGAAGATC	TGTGCCACCA	TCATCAGGCT	3900
GGGTTTTAAA	AGATTATGGT	TTATATTTAA	TGTGTATGAG	TGTTTTGTTT	GCATGTATAT	3960
CTGTACATGA	CAGGTGTGCC	TGGTGTTTAC	AGAGGCCAGA	AGAACATACC	AGATCCCCCT	4020
GGAAGTGAAG	TTACAGACAG	TCGTGAGCCA	TCTCGGGGTT	GCTGGGGACA	GAATCCGAGG	4080
GCTCTTCTTG	AGTAGCAAGT	GCTTCTAACC	GCTTAGGCCT	CTCTGCAGCC	CCCACTTACA	4140
GGATTTAAAG	GTAGAACAAG	GTTTGTCAAC	TGTCCTGGAG	ACCCTGGCCT	TTAATTCCAG	4200
AACTCTGGAG	GTAGAGACAG	ATGATTCTCT	ATGAAGTTCA	GGCGAGCCTG	GTCTACACAG	4260
AGTGCCGCAT	GATAGCAAGA	AGAAGATCCT	GTCTTTAAAA	GAGACGAGAG	GGGTTGGGGA	4320
TTTAGCTCAG	TGGTAGAGCG	CTTGCCTAGC	AAGCACAAGG	CCCTGGGTTC	GGTCCCCAGC	4380
TCCGAAGAAA	AAAAAGAGAC	GAGAGCCAGT	GGTTGGTGCA	CGTCTTTGAT	CCCAGTACTC	4440
TGGAGGCAGA	GGTAGTGGAT	CTCTCTTGAG	TTCAAGGACA	GCGTGGTCTA	CAAAGTGAGT	4500
TTTAGGACAT	CCAGGATTAC	ACGCACAGAA	ACCTTGCTCTC	ATAAAACAAC	AAACAAGACA	4560
AGACAGAAAC	TCTCCTAACG	TAGACCGCCA	CACCTGATTT	TTAAAAGCTC	TCAGTGAAAC	4620
TGAGCATGGT	AGCACATGTT	TGTAATCCCA	GCAGACATGT	GGGGAGACAA	AGGAATGGAC	4680
TCAGACTCAG	CCGGAGAGCA	AGTTCACGGC	TAGACTGGAC	CATTCCTACA	ATGAGGTAGG	4740
AATTGGGGTT	AGCACATCAA	GTAAGTAACC	CTGGAAAACAA	GTTTGACTTG	TCCAAGGTCA	4800
CACAGCAATG	TCTGGAAAGC	TAAGTCTGGT	TCCAAGGCCC	CCCCTTCCTC	CCTCTCTCCC	4860
TCTCTATAAT	TGAAAAGTCC	ACTGCTTGGC	AAAAACTCCC	AGGACTATAT	TAAACACAAA	4920
TGCTGGTGT	CTCCATGTCT	TAGGGCTTTT	ATCCTAGAAG	GAATTCAAAC	ACACAACACG	4980
AATACCCAC	AGAAAGGAGG	GCAGGGTGGA	GGGGTAAGGG	AGAGAGGAGG	AACTTCAGGC	5040
TACTGGGGGT	ATTAACCAGC	TCTGTACCCC	ATCCACACAG	ACCCAAGTTA	GAAAAGAGCA	5100
GGAGAGGGGG	TCTGGAGAGG	TTGTTAACTG	GCCCAGCAGT	TTGGCCTGCT	CTTGCAGGGG	5160
CCCAGCTCTG	TTCCCAGCAC	CCATTTCACT	GGCTCACAAC	TTTAACTCC	AGCCCCAAGG	5220
ACTCTGCTTC	CCTCTGAGAG	CTCTGTACTT	AACAGGGACA	CACAGACACA	TACAATTAAA	5280
AAAATGTTTT	AAAGTGAGAG	ACGCTCTAGA	CAGGCTAGCA	AGTATTGAGT	TGTGGCAGGT	5340
ACAGCTATTT	TAATAGTGAT	TTCAGGTTAG	AACCCTGGGG	AGGGGGAACC	AGGAGTTAAA	5400
CTATGTTAAA	TCAGAAAGAC	CCAAAGCCAA	TCTGGTGGA	GCTGCCATTG	GAGGTTCTAA	5460
CAAGTCTGGC	TTGTCAGGGA	AAGGCTCAGA	ATGAAGGTTT	GAGCTGGGGC	ATCATTAGTG	5520

TATAAAAAGT ATGAAAACAC TCTAGGAAGA AGACAAGAGG AGGAACACCA CGGAGAGCGA	5580
GCCTTACGAT GTTCCAGCAC GTAGACGCCA AAGTGAAGCC AGGAAACCAA GCACAGGGAC	5640
CAGGAAAGCC CAAAGTTCAT TGTGAAAAGG ACAAAGGCTT CATCCTGGGA AACTAGGCTG	5700
GGAGAGGCCG TGTAAATAA AGACAGACAC ACCCATCAAA ATGACCCACA GAGGGCTTCA	5760
TGATTACAAT AGTATTTTCAT AGGCGGATTT GGGCAGAAAT CTGAATGCAG GGGATTACAG	5820
AGTAAATGCT GACTTTTGGG TAAGAATGGC AGATCACAGG ACAGGTGTGT GACTCACATC	5880
TTTAAAGCAC ACTCCCAGGG CAGAGGTAGT GAGTTTGAGT TTAGGGGCTA GTCTGGTCTG	5940
GACTGGAAAT TCTATGAGAC CCTGTTCTCA AAAACTAAAG TATTTGGGAA AAAAGAACTT	6000
CTGAGGGAAA TGGAGGCCGT GTAGGTCTCT CTGGGAGCCC GTGCGGCAGG TGGCGAGGGA	6060
GGATCTGAAA TGGGGAGAGT CAGCAGACTG CTGGACCTTT CCTAGCCAGC AGAGATGCTA	6120
AGGCAGGTGA AGATTAGGTC TCATGGACCT GACACCCGTG CACACAGGCA GCATGGCGCC	6180
TTCAAAGCTC TAGTGGATGT GATTGCCCCA GACAAGTCTG CCCCAAAGCT CATCTTCGTC	6240
CATTAATAGA AAAAAGGTTT CTTCTGACCA AGGAAGCTGT TCTCTCTGGA AAACAATCAC	6300
TTAACAAGGA CATTACTAAC ACGAAGCTGC TGTCCGATCA CATCACCATG ACGCAAGCAC	6360
TTCCCTTGGG GTTCATACGC AGTGACTCAG TGCTCACGAC CCTGTGCTAG GCTTGGCCCT	6420
CACTCCTTTT CCGCTGGAAT TAAGTGGGGA GTCAGACACC CCAGAGGACC TGCCCAAGCC	6480
AGAAAGCTTC AAGCCACAGG AGCCAGTGTG TCCTTGGCTT CCCTACACAT GAGCTGTCTC	6540
TTATCCTCGA TCGAGGGCCT CACAGTCATT CCTGAAAAGA TCTGGCCCCC AGCCCTGAGT	6600
ATGGAAGGCT AACTTGGCTA CCAGTCCCCA CTGTCCTTAT TAGGAAGAGG CAAAACCGTC	6660
CTCTGGCACT CTCTTGAAGC ATACTGGTAT ATCCGAGAGA GGTAACAGGA GCCGATGGGA	6720
GCTGGGAGGG TCCTGGCCTA GGCATAGTCT AGAAGACTTG GGCTAAGTAG TCTGGGTCCC	6780
CAAACCATAA CATTTTTCTG GTGACTAAAG AAAAGGAGTC TGTAAGCCTA AAGCAGAATG	6840
TGGTGATACA CGCCTACAGT CCTAGCACTG GAGAGGTGGA GATAGAAAGA TCAAGAGTTC	6900
AATGCCAGCT TTCTGCTATG TAGTAAGGTC AAGGTCAGCC TGGACTAAAC GACTGCCTTA	6960
GAAACAACCA AATGACTTAC CGTCTAAAGT CAGGAACTAC ACTTGCTTTC TCAGACTGTG	7020
TCTGTCTGTC TGGGGCTCCT CCCATTTCTT CTCCTAACAA CATCCACTTC CACTCCTGCC	7080
TTAGATCTGA GATAGTACCA GCCTCAGGGC ATGGGGTCTC CCCATAGCTT TTCTCTGCA	7140
GTACTGTGGG CTCACCTAGG ACTGTTTCTG AACTATATCC TACCCTAGCT CTCTACCCTA	7200
GAAGGCCTGA AACTCACAGA AATTCTCCTG CCTCTGCTTT CCAATGGCTG GGGTTAAAAG	7260
CATGTGTCAC AACTGTCCTT TTTATTCTTT TAATATCGAG ACAGGGTCTC ACCAAGTTGC	7320
CCCAAGACGC CAGCCACACC TGGGACAGGG CAGGCCTTTG GCTCTATGTT CAGTCTTGAC	7380

TCCATGACTG TGGCCGCTAG CCCATGAGGC TGCGCGTGGG AATTTCTTC TGAAAGCTCA	7440
CCTGGTATCG ATGCTTCCTC TTATCCTACA CCACAATAA CAAACCTGCC CCACCTCCTG	7500
GTCCTGACCC TGCTGCAGAC CTGCTAGTCC TTGGTGAATG AGACCTGGGG ACCCCTCTAG	7560
TCTGTTGAGA GCTGCTGAAA TGCTCAACTA TGATTTCAG GTGACCCTCA AGTCGGCTCA	7620
CCTCCCTGAT TGCACAGCAC CAATCACTGT GGCCTGGCT CCCGTCACAC GGTGGCCAGT	7680
GACAGCCTGA TGGCTGGCTC CCCTCCTCCA CCACCCTCTG CATTGACAGG CCCACGTGTG	7740
TCCCCAGATG CCTGAATCAC TGCTGACAGC TTGGGACCTG TCAGCTGTGG GCTCCTGGGG	7800
AGCCACTGGG GAGGGGGTTA GCAGCCACGC TGTCGCCTCC TAGCCAACAC CTGCAGACAT	7860
AAATAGACAG CCCAGCCCGC TCAGGCAGCA GAGCAGAGCT GCACGACGCG TCGATCCCAA	7920
GGCCCAACTC CCCGAACCAC TCAGGGTCCT GTGGACAGCT CACCTAGCTG CAATGGCTAC	7980
AGGTAAGCGC CCCTAAAATC CCTTTGGCAC AATGTGTCTT GAGGGGAGAG GCAGCGACCT	8040
GTAGATGGGA CGGGGGCACT AACCTCAGG GTTTGGGGTT CTGAATGTGA GTATCGCCAT	8100
CTAAGCCCAG TATTTGGCCA ATCTCAGAAA GCTCCTGGCT CCCTGGAGGA TGGAGAGAGA	8160
AAAACAAACA GCTCCTGGAG CAGGGAGAGT GTTGGCCTCT TGCTCTCCGG CTCCCTCTGT	8220
TGCCCTCTGG TTTCTCCCCA GGCTCCCGGA CGTCCCTGCT CCTGGCTTTT GGCCTGCTCT	8280
GCCTGCCCTG GCTTCAAGAG GGCAGTGCCT TCCAACCAT TCCCTTATCC AGGCTTTTTG	8340
ACAACGCTAT GCTCCGCGCC CATCGTCTGC ACCAGCTGGC CTTTGACACC TACCAGGAGT	8400
TTGTAAGCTC TTGGGGAATG GGTGCGCATC AGGGGTGGCA GGAAGGGGTG ACTTTCCCCC	8460
GCTGGAAATA AGAGGAGGAG ACTAAGGAGC TCAGGGTTTT TCCCGACCGC GAAAATGCAG	8520
GCAGATGAGC ACACGCTGAG CTAGGTTCCT AGAAAAGTAA AATGGGAGCA GGTCTCAGCT	8580
CAGACCTTGG TGGGCGGTCC TTCTCCTAGG AAGAAGCCTA TATCCCAAAG GAACAGAAGT	8640
ATTCATTCTT GCAGAACCCC CAGACCTCCC TCTGTTTCTC AGAGTCTATT CCGACACCCT	8700
CCAACAGGGA GGAAACACAA CAGAAATCCG TGAGTGGATG CTTTCTCCCC AGGCGGGGAT	8760
GGGGGAGACC TGTAGTCAGA GGGGGGGG AGCACAGCCA ATGCCCCGTCC TTGCCCCCTGC	8820
AGAACCTAGA GCTGCTCCGC ATCTCCCTGC TGCTCATCCA GTCGTGGCTG GAGCCCCGTGC	8880
AGTTCCTCAG GAGTGTCTTC GCCAACAGCC TGGTGTACGG CGCCTCTGAC AGCAACGTCT	8940
ATGACCTCCT AAAGGACCTA GAGGAAGGCA TCCAAACGCT GATGGGGGTG AGGGTGGCGC	9000
CAGGGGTCCC CAATCCTGGA GGGGACTGA CTTTGAGAGA CTGTGTTAGA GAAACACTGG	9060
CTGCCCTCTT TTTAGCAGTC AGGCCCTGAC CCAAGAGAAC TCACCTTATT CTTCAATTCC	9120
CCTCGTGAAT CCTCCAGGCC TTTCTCTACA CTGAAGGGGA GGGAGGAAAA TGAATGAATG	9180

AGAAAGGGAG	GGAACAGTAC	CCAAGCGCTT	GGCCTCTCCT	TCTCTTCCTT	CACTTTGCAG	9240
AGGCTGGAAG	ATGGCAGCCC	CCGGACTGGG	CAGATCTTCA	AGCAGACCTA	CAGCAAGTTC	9300
GACACAAACT	CACACAACGA	TGACGCACTA	CTCAAGAACT	ACGGGCTGCT	CTACTGCTTC	9360
AGGAAGGACA	TGGACAAGGT	CGAGACATTC	CTGCGCATCG	TGCAGTGCCG	CTCTGTGGAG	9420
GGCAGCTGTG	GCTTCTAGCT	GCCCCGGTGG	CATCCCTGTG	ACCCCTCCCC	AGTGCCTCTC	9480
CTGGCCCTGG	AAGTTGCCAC	TCCAGTGCCC	ACCAGCCTTG	TCCTAATAAA	ATTAAGTTGC	9540
ATCATTTTGT	CTGACTAGGT	GTCCTTCTAT	AATGACGCGT	CGTGCCCAAC	TATGCTCGCC	9600
ATGATGCTCA	ACACTACGCT	CTCTGCTTGC	TTCCTGAGCC	TGCTGGCCCT	CACCTCTGCC	9660
TGCTACTTCC	AGAACTGCCC	AAGAGGAGGC	AAGAGGGCCA	CATCCGACAT	GGAGCTGAGA	9720
CAGGTACCAC	TGTGGTCCGT	TCAGGGCTGC	TGACAGTGCC	GTAGGAAGGG	TCATGGGCTA	9780
GGAGAGAGGG	AAACCTTGTC	TGAGCAGTCA	GACTTTAGGG	GAGGTTCTTG	GAAGGAAGCA	9840
GTTATCTTAT	ATGGAGTAGA	TGGGTTTCCC	AGAACGGTAA	GAGGGGACCA	GGTGCCAGAG	9900
AAGCCACATA	AAGGACAGTG	TCCCCAGGCA	GGGGATATGC	CAGAAAATGA	GAGATACTTA	9960
TCACTGGGCT	TGGGATGAGA	ACGGGTAAAA	CTGGGTACCC	TGGCCTCCTC	TGCACAGCTG	10020
GAGGTGGCCG	GTGGTATGTT	GGCTCACCAG	GACTGGGTAG	ATGGTACGAA	ACTGTTCTCG	10080
CCTGAGTACA	AAGCCTTTCC	CACCCAGCTC	AAACTCTCTT	AGCTCCTTTT	TTAGCCAGCT	10140
GCACCGGTTT	CTTCCTGTCC	ACGGAAGACG	GCCATTGCCC	TGTGTCTGAG	CGGAGTATGT	10200
CCCACATCTA	GCCTCAGCCT	CGTGCCCAGA	TCTGCTGTAC	TGTATGTTCA	GCTCTGAGTC	10260
TGCCCTTCCG	GCAGGGCTGA	AGGGAATCCA	GTCAGTAGGC	TCAAATCTGG	TCAGGTCACA	10320
GGTGGCTCAG	TTTTGAACAA	GCTCGATGGG	CAGTAGGCAG	TTCACCGAGT	CTGCCTTCCG	10380
TTTGCTGAGT	TCCTTTGGAG	ACTTCCGAGG	CACTAGGTGT	GTCTTGCAAC	CATCAGCCTA	10440
ATTCGGTCCT	TGCCACCTTC	CTACTAGGGC	ATAATAGGTT	GGCGGGAGGT	AAAAGCCCAC	10500
CAGCGTGGGG	CAGGGGTAAG	AGTGAGCGAG	CCGTAGGTAC	AGGAAAGAGG	ATCTTGGAAT	10560
GTGTAGGGCC	ATCTGAATGT	CGGAGAGGTA	AGTCTCTGAG	AGACTGCTGC	ACACCGGTGA	10620
CACATCAGAG	CTGAGGAGGT	CCCCCAAGTG	TTGTCTCCCC	CGCCCCCGC	CCCATACGAC	10680
TCTGTCAAAG	CAGGAGAGGG	TTTTGAGACC	TCATGAGAAC	TGATCCTCCT	GATAACCTAG	10740
CCGGTTAGAT	TTCCACTCTC	GCCCTTTACG	GCTGCTTCGT	CCTAGATAGA	GCCAGAGCAT	10800
CTGGCCGGTG	AAGCTGGGAT	AGCAGCAGGG	TGACCTTAGG	TTCCCAACGC	CCCTCTTGGC	10860
CTGGCTCCAG	CTGACCCGCG	TCCTTCCCCG	CAGTGTCTCC	CCTGCGGCCC	TGGCGGCAAA	10920
GGGCGCTGCT	TCGGGCCGAG	CATCTGCTGC	GCGGACGAGC	TGGGCTGCTT	CCTGGGCACC	10980
GCCGAGGCGC	TGCGCTGCCA	GGAGGAGAAC	TACCTGCCCT	CGCCCTGCCA	GTCTGGCCAG	11040

AAGCCTTGCG	GAAGCGGAGG	CCGCTGCGCT	GCCGCGGGCA	TCTGCTGCAG	CGATGGTGCG	11100
CACAAAGCCA	GGCGGGCTGA	GCATGGGGAA	TGGATGGGGT	GGGTGGGAGG	TAAAGGGGGG	11160
CTAAGTGGGG	GA CTGAGGAA	TCAGGACCGG	AGATGGAGGG	TGAGTAGTAT	GAAGGGGGTC	11220
GAGAGTTGGA	ACGTAGCAGG	G TAGGATAAA	GGGATTGTG	GGGATGGCGC	CCCTATAGGT	11280
GCGCCACCC	CAGGACGCCT	GACCTCACAC	AGCCCTTCCT	TCAGAGAGCT	GCGTGGCCGA	11340
GCCCCAGTGT	CGAGAGGGTT	TTTTCCGCCT	CACCCGCGCT	CGGGAGCAGA	GCAACGCCAC	11400
GCAGCTGGAC	GGGCCAGCCC	GGGAGCTGCT	GCTTAGGCTG	GTACAGCTGG	CTGGGACACA	11460
AGAGTCCGTG	GATTCTGCCA	AGCCCCGGGT	CTACTGAGCC	ATCGCCCCCC	ACGCCTCCCC	11520
CCTACAGCAT	GGAAAATAAA	CTTTTAAAAA	ATGCACCCTG	GTGTCTGTCT	CTCTTTCTGG	11580
GGTGGGGAGA	AAAGGGGGGA	GAGGAATTGG	AGTGGGAACT	TTCTACTCTG	CTCTGACTGA	11640
TCCCCACATC	CAAAGTCGTG	CATAAGATAC	GCCCCACCG	CCAGAAGGGG	CAGAACCTAT	11700
AAGTCTTAGA	GTATAAAGGA	AGCTTCTGCT	GCTCCTGGAT	ACCCACATAA	TACTCAGAAA	11760
AAAAGGCAAG	TCAGAAGAAG	GGAAAGATCT	GAGATCCAGA	GGAGCCTGAA	GGGTCAGGGT	11820
GACTTAGCAA	GTTTCTATCT	GAGACCGAAA	TAAAAGGACA	TTGTGGACAA	GAGAAACAGA	11880
GCAGGACATG	AGGAGAGACA	GGATCAGCAA	GAGTGACAGA	GAAAGAGGGG	ACAGGCCAGG	11940
GGTGGCCATC	TCAGCCCTGA	TTTCACCCAG	ACTAAGGCAA	AAACAACGTG	AAGGACTCTT	12000
AACCAAGGCT	GTGCTTGGAT	GGGAGGAGAA	GGTACAGAGA	CATTACCCCA	GACCTAAAGA	12060
AGACAATGCC	ACCCGCCTTC	TCTCCAGGTG	CTCCACCATC	AAGACCCAGC	CACTGAGAGG	12120
CAGACTCCAG	TAAGAGTCCA	GCTACAAGTC	CTCTACAGGC	ACATGTTCAA	ACCGTCACAC	12180
CCCACTCAG	GCAGGGAAAT	AGACAAGATA	GGCTGGAGTT	GTGGCTCAGG	AGCAGAAGTC	12240
TTCCCTAGCT	ATGGTTCCAG	CACCATGGAG	ACGGAAGGGG	GA CTGAGCGG	GGGGGGGGGC	12300
GGGGGGGAGG	AAAAAGTAGC	AGCTACTAGG	GGCATTTCTA	TGACCCTTGT	CCTCAACCAT	12360
AGCTAGAGAC	CCAGAGGAAC	ACAGAAGTCC	AGCAGCAAGG	CGCACATGCT	TGCAATAGCT	12420
CCCAGACGTA	AATACTTCAT	TCCGTTCCGC	ACATCCGGGT	CATCAGCACT	TGACTCCCCC	12480
CCCCACACTT	CTTATTACCT	CCTCTTTTTT	CTAAAATTTT	AGATTTATTC	ACTTATGTAG	12540
ATGGGTGTTT	TTGGTTTTGT	ATGAATGTCT	GAGCACCATG	TGTGTGCCTG	GCGCCTCAGA	12600
GGTCAGGAGA	GGGCATCGGG	TCCCCTGGAA	CTAGTTACAG	GTGGTTACAG	CCTACCGTGT	12660
GGGCACTAGG	AACTGAATCC	CAGTCCTCTT	AACTGACCCG	CACATCCAAA	CCCAGGCTTC	12720
AGCCCTCAT	CAGCCTGTCC	CTCCTCCAGG	CCCTCAGGTG	TCTCCCGTCT	CCGGCTGCTC	12780
TCCCAGACAT	CCTTCCATCC	TCTGGTCTCC	CTGCTCCTCG	CCCTCCTGTT	AACATCCTTT	12840

CTCTCTGCCC	CATCTGTCTT	GGGCATCCTC	TCCTGCGAGC	TGCAGCAAGG	TCAGGATGGT	12900
TTACCTCATT	TGGGATGGCC	TGCAGGTTCT	GAGGTCAGGG	GCAACTACAG	AGAAGAGAGA	12960
GATTAGTCTG	ATTGACTTAA	GGTGTTTCAG	CAAGGTCAGC	TCTGCCCAGA	CTCACGGTCT	13020
TTTACCCAGA	TGCCAGCTCT	CTTCCCATCT	CCTCGGTGCC	TATACACCTC	TCTGCATGCC	13080
CCGGTTTAGA	CAGGTAGCAC	AGGGGCCAGG	CAGACTCCTA	TCCCAGCCTC	CTCCTTCTGT	13140
GGCCCTCTTA	GGGTCTGACC	TCCAATAGGG	CAGGGCCAGG	GAAGGGCCAG	ACCAAAAAGG	13200
GACAGAAAGA	AGCGTGCGAG	GCGGCATGGG	CACACTTGAT	TCAACCCCTA	CGGCTGGTGT	13260
ATGGGCAGCT	TTAGAATGAA	GGTCAGATTC	TCACTTCGAG	CCTCTGCGCA	GGTGGAGTGT	13320
TGTAAGCGTC	TCGCTTTCTT	CCACCTGTTT	CTGGAAGAAT	CAGGCTCCTC	TTCCTCGAGG	13380
AGAGAATTAT	ACCTGCTCAC	CCTACTTCTG	CCTACTGGAC	ATAATATATA	TTTTTTTCTT	13440
TTCAGGGAGT	CCTTTCTCTA	GCTACAGAGC	CATTTAAGGG	CACTCCCAGA	GTTACAGCA	13500
GATGCTTGCC	TCCTCTCTTC	AGCCTCCAGA	AGCAGAGAGG	CTTGTGAGCA	AATGCCAGGA	13560
CCTCTGACCT	CCACACAGAC	GCTGTGCTGT	GTGCACAGCC	CTCAAGCACA	CAGCGAAGCA	13620
ATAGTGAAAA	GTAACCTAGA	CCATTTTCAG	GCTGGGGAGA	TGGCTCAGGA	GATAAGAGAT	13680
CACTGCTCAA	CTTGAGCCTC	GGGACCACAG	GTAAGACCAA	CTTGTCTGCT	GCAAGAGAGC	13740
TGCCTGGTGA	GATTGGGACA	CACAGAGGCA	GAGTTCATCT	AGGACCGGGC	ACGTCCTGTG	13800
TTTGCCGAGG	TCCCACACCC	GCGGATCCCG	GCCCGCAGCA	GCTCTCTGCT	CCCAGAACCC	13860
GTGAGAAAGA	GACCTCACCG	CCTGGTCAGG	TGGGCACTCC	TGAGGCTGCA	GAGCGGAAGA	13920
GACCACCAAC	ACTGCCCACC	CCTGCCCCACA	TCCCTGGCCC	AAGAGGAAAC	TGTATAAGGC	13980
CTCTGGGTTT	CGTGGGGGAG	GGCCAGGAG	CGTCAGGACC	CCTGCCTGAG	ACACCGCCGG	14040
AACCTGAGGG	AAACAGACCG	GATAAACAGT	TCTCTGCACC	CAAATCCCAT	GGGAGGGAGA	14100
GCTGAACCTT	CAGAGAGGCA	CACAAGCCTT	GGAAACCAGA	AGAGACTGCT	CTCTGTACAT	14160
ACATCTCGGA	CGCCAGAGGA	AAACACCAAA	GGCCATCTGG	AACCCTGGTG	CACTGAAGCT	14220
CCTGGAAGGG	GCGGCACAGG	TCTTCTGGT	TGCTGCCGCC	ACAGAGAGCC	CTTGGGCAGC	14280
ACCCCGCCTG	GTGAACTCAA	GACACAGGCC	CACAGGAACA	GCTGAAGACC	TGCAGAGAGG	14340
AAAACTACA	CGCCCGAAAG	CAGAACACTC	TGTCCCCATA	ACGGACTGAA	AGAGAGGAAA	14400
ACAGGTCTAC	AGCACTCCTG	ACACACAGGC	TTATAGGACA	GTCTAACCAC	TGTCAGAAAT	14460
AGCAGAACAA	AGTAACACTA	GAGATAATCT	GATGGTGAGA	GGCAAGCGCA	GGAACCCAAG	14520
CAACAGAAAC	CAAGACTACA	TGGCATCATC	GGAGCCCAAT	TCTCCCACCA	AAACAAACAT	14580
GGAATATCCA	AACACACCAG	AAAAGCAAGA	TCTAGTTTCA	AAATCATATT	TGATCATGAT	14640
GCTGCAGGAC	TTCAAGAAAG	ACGTGAAGAA	CTCCCTTAGA	GAACAAGTAG	AAGCCTACAG	14700

AGAGGAATCG	CAAAAATCCC	TGAAAGAATT	CCAGGAAAAC	ACAATCAAAC	AGTTGAAGGA	14760
ATTAAAAATG	GAAATAGAAG	CAATCAAGAA	AGAACACATG	GAAACAACCC	TGGATATAGA	14820
AAACCAAAGG	AAGAGACAAG	GAGCCGTAGA	TACAAGCATC	ACCAACAGAA	TACAAGAGAT	14880
GGAAGAGAGA	ATCTCAAGAG	CAGAAGATTC	CATAGAAATC	ATTGACTCAA	CTGTCAAAGA	14940
TAATGTAAAG	CGGAAAAAGC	TACTGGTCCA	AAACATACAG	GAAATCCAGG	ACTCAATGAG	15000
AAGATCAAAC	CTAAGGATAA	TAGGTATAGA	AGAGAGTGAA	GACTCCCAGC	TCAAAGGACC	15060
AGTAAATGTC	TTCAACAAAA	TCATAGAAGA	AAACTTCCCT	AACCTAAAAA	AAGAGATACC	15120
CATAGGCATA	CAAGAAGCCT	ACAGAACTCC	AAATAGATTG	GACCAGAAAA	GAAACACCTC	15180
CCGTCACATA	ATAGTCAAAA	CACCAAACGC	ACAAAATAAA	GAAAGAATAT	TAAAAGCAGT	15240
AAGGGAAAAA	GGTCAAGTAA	CATATAAAGG	CAGACCTATC	AGAATCACAC	CAGACTTCTC	15300
GCCAGAAACT	ATGAAGGCCA	GAAGATCCTG	GACAGATGTC	ATACAGACCC	TAAGAGAACA	15360
CAAATGCCAG	CCCAGGTTAC	TGTATCCTGC	AAAACCTCTCA	ATTAACATAG	ATGGAGAAAC	15420
CAAGATATTC	CATGACAAAA	CCAAATTTAC	ACAATATCTT	TCTACAAATC	CAGCACTACA	15480
AAGGATAATA	AATGGTAAAG	CCCAACATAA	GGAGGCAAGC	TATACCCTAG	AAGAAGCAAG	15540
AAACTAATCG	TCTTGGCAAC	AAAACAAAGC	GAATGAAAGC	ACACAAACAT	AACCTCACAT	15600
CCAAATATGA	ATATAACGGG	AAGCAATAAT	CACTATTCCCT	TAATATCTCT	CAACATAAAT	15660
GGCCTTAACT	CCCCAATAAA	AAGACATAGA	TTAACAAACT	GGATACGCAA	CGAGGACCCT	15720
GCATTCTGCT	GCCTACAGGA	AACACACCTC	AGAGACAAAG	ACAGACATTA	CCTCAGAGTG	15780
AAAGGCTGGA	AAACAATTTT	CCAAGCAAAT	GGTCAGAAGA	AGCAAGCTGG	AGTAGCCATT	15840
CTAATATCAA	ATAAAATCAA	TTTTTAACTA	AAAGTCATCA	AAAAAGATAA	GGAAGGACAC	15900
TTCATATTCA	TCAAAGGAAA	AATCCACCAA	GATGAACTCT	CAATCCTAAA	TATCTATGCC	15960
CCAAATACAA	GGGCACCTAC	ATATGTAAAA	GAAACCTTAC	TAAAGCTCAA	AACACACATT	16020
GCACCTCACA	CAATAATAGT	GGGAGATTTT	AACACCCCAC	TCTCATCAAT	GGACAGATCA	16080
TGGAAACAGA	AATTAAACAG	AGATGTAGAC	AGACTAAGAG	AAGTCATGAG	CCAAAGGGAC	16140
TTAACGGATA	TTTATAGAAC	ATTCTATCCT	AAAGCAAAAAG	GATATACCTT	CTTCTCAGCT	16200
CCTCATGGTA	CTTTCTCCAA	AATTGACCAT	ATAATTGGTC	AAAAACGGG	CCTCAACAGG	16260
TACAGAAAGA	TAGAAATAAT	CCCATGCATG	CTATCGGACC	ACCACGGCCT	AAAACCTGGT	16320
TTCAATAACA	ATCAAGGAAG	AATGCCCAT	TATACTTGGA	AACTGAACAA	TGCTCTACTC	16380
AATGATAACC	TGGTCAAGGA	AGAAATAAAG	AAAGAAATTA	AAAACTTTTT	AGAATTTAAT	16440
GAAAATGAAG	GTACAACATA	CCCAAACCTA	TGGGACACAA	TGAAAGCTGT	GCTAAGAGGA	16500

AAACTCATAG	CGCTGAGTGC	CTGCAGAAAG	AAACAGGAAA	GAGCATATGT	CAGCAGCTTG	16560
ACAGCACACC	TAAAAGCTCT	AGAACAAAAA	GAAGCAAATA	CACCCAGGAG	GAGTAGAAGG	16620
CAGGAAATAA	TCAAACCTCAG	AGCTGAAATC	AACCAAGTAG	AAACAAAAGG	ACCATAGAAA	16680
GAATCAACAG	AACCAAAGT	TGGTTCTTTG	AGAAAATCAA	CAAGATAGAT	AAACCCTTAG	16740
CCAGACTAAT	GAGAGGACAC	AGAGAGTGTG	TCCAAATTAA	CAAAATCAGA	AATGAAAAGG	16800
GAGACATAAC	AACAGATTCA	GAGGAAATTC	AAAAAATCAT	CAGATCTTAC	TATAAAAACC	16860
TATATTCAAC	AAAACCTGAA	AATCTTCAGG	AAATGGACAA	TTTCTAGAC	AGATACCAGG	16920
TACCGAAGTT	AAATCAGGAA	CAGATAAACC	AGTTAAACAA	CCCCATAACT	CCTAAGGAAA	16980
TAGAAGCAGT	CATTAAAGGT	CTCCCAACCA	AAAAGAGCCC	AGGTCCAGAC	GGGTTTAGTG	17040
CAGAATTCTA	TCAAACCTTC	ATAGAAGACC	TCATACCAAT	ATTATCCAAA	CTATTCCACA	17100
AAATTGAAAC	AGATGGATCA	CTACCGAATA	CCTTCTACGA	AGCCACAATT	ACTCTTATAC	17160
CTAAAAACA	CAAAGACACA	ACAAAGAAAG	AGAACTTCAG	ACCAATTTCC	CTTATGAATA	17220
TCGACGCAA	AATACTCAAC	AAAATTCTGG	CAAACCGAAT	CCAAGAGCAC	ATCAAAACAA	17280
TCATCCACCA	TGACCAAGTA	GGCTTCATCC	CAGGCATGCA	GGGATGGTTT	AATATACGGA	17340
AAACCATCAA	CGTGATCCAT	TATATAAACA	AACTGAAAGA	ACAAAACCAC	ATGATCATTT	17400
CATTAGACGC	TGAGAAAGCA	TTTGACAAAA	TTCAACACCC	CTTCATGATA	AAAGTCCTGG	17460
AAAGAATTGG	AATTCAAGGC	CCATACCTGA	ACATAGTAAA	AGCCATATAC	AGCAAACCAG	17520
TTGCTAACAT	TAAACTAAAT	GGAGAGAAAC	TTGAAGCAAT	CCCACTAAAA	TCAGGGACTA	17580
GACAAGGCTG	CCCACTCTCT	CCCTACTTAT	TCAATATAGT	TCTTGAAGTT	CTGGCCAGAG	17640
CAATCAGACA	ACAAAAGGAG	GTCAAGGGGA	TACAGATCGG	AAAAGAAGAA	GTCAAAATAT	17700
CACTATTTGC	AGATGATATG	ATAGTATATT	TAAGTGATCC	CAAACATTCC	ACCAGAGAAC	17760
TACTAAAGCT	GATAGACAAC	TTCAGCAAAG	TGGCTAGGTA	TAAAATTAAC	TCAAATAAAT	17820
CAGTTGCCTT	CCTCTATACA	AAAGAGAAAC	AAGCCGAGAA	AGAAATTAGG	GAAACGACAC	17880
CCTTCATAAT	AGACCCAAAT	AATATAAAGT	ACCTCGGTGT	GACTTTAACA	AAGCAAGTAA	17940
AAGATCTGTA	CAATAAGAAC	TTCAAGACAC	TGAAGAAGGA	AATTGAAGAA	GACCTCAGAA	18000
GATGGAAAGA	TCTCCCGTGC	TCATGGATTG	GCAGGATTAA	TATAGTAAAA	ATGGCCATTT	18060
TACCAAAGC	AATCTACAGA	TTCAATGCAA	TCCCCATCAA	AATACCAATC	CAATTCTTCA	18120
AAGAGTTAGA	CAGAACAATT	TGCAAAATTC	TCTGGAATAA	CAAAAAACCC	AGGATAGCTA	18180
AAGCTATCCT	CAACAATAAA	AGGACTTCAG	GGGGAATCAC	TATCCCTGAA	CTCAAGCATG	18240
ATTACAGAGC	AATAGTGATA	AAAACGTCAT	GGTATTGGTA	CAGAGACAGA	CAGATAGACC	18300
AATGGAATAG	AATTGAAGAC	CCAGAAATGA	ACCCACACAC	CTATGGTCAC	TTGATTTTTG	18360

ACAAAGGAGC CAAAACCATC AAATGGAAAA AAGATAGCAT TTTCAGCAAA TGGTGCTAGT	18420
TCAACTGGAG GTCAACATGT AGAAGAATGA AGATCGATCC ATGCTTGTCA CCCTGTACAA	18480
GCTTAAGTCC AAGTGGATCA AGGACCTCCA CATCAAACCA GACACACTCA AACTAATAGA	18540
AGAAAACTA GGAAGCATC TGGAACACAT GGGCACTGGA AAAAATTTCC TAAACAAAAC	18600
ACCATGGCTT ACGCTCTAAG ATCAAGAATC GACAAATGGG ATCTCATAAA ACTGCAAAGC	18660
AACTGTAAGG CAAAGGACAC TGTGGTTAGG ACAAAACGGC AACCAACAGA TTGGGAAAAT	18720
ATCTTTACCA ATCCTACAAC AGATAGAGGC CTTATATCCA AAATATACAA AGAACTCAAG	18780
AAGTTAGACC GCAGGGAAAC AAATAACCCT ATTAAAAAAT GGGGTTTCTG GCTAAACAAA	18840
GAATTCACAG CTGAGGAATG CCAAATGGCT GAGAAACACC TAAAGAAATG TTCAACATCT	18900
TTAGTCATAA GGGAAATGCA AATCAAAACA ACCGTGAGAT TTCACCTCAC ACCAGTGAGA	18960
ATGGCTATGA TCAAAACTC AGGGGACAAC AGATGCTGGC GAGGATGTGG AGAAAGAGGA	19020
ACACTCCTCC ATTGTTGGTG GGATTGCAAA CTGGTACAAC CATTCTGGAA ATCAGTCTGG	19080
AGGTTCTCTA GAAAATTGGA CATTGAACTG CCTGAGGATC CAGCTATACC TCTCTTGGGC	19140
ATATACCCAA AAGATGCCCC AACATATAAA AAAGACACGT GCTCCACTAT GTTCATTGCA	19200
GCCTTATTTA TAATAGCCAG AAGCTGGAAA GAACCCAGAT GCCCTTCAAC AGAGGAATGG	19260
ATACAGAAAA TGTGGTACAT GTACACAATG GAATATTACT CAGCTATCAA AAACAACGAG	19320
TTTATGAAAT TCGTAGGCAA ATGGTTGGAA CTGGAAAATA TCATCCTGAG TAAGCTAACC	19380
CAATCACAGA AAGACATACA TGGTATGCAC TCATTGATAA GTGGCTATTA GCCCAAATGC	19440
TTGAATTACC CTAGATACCT AGAACAAATG AAACCTCAAGA CGGATGATCA AAATGTGAAT	19500
GCTTCACTCC TTCTTTAAAA GGGGAACAAG AATACCCTTC GCAGGGAAGA GAGAGGCAAA	19560
GATTAAACA GAGAATGAAG GAACACCCAT TCAGAGCTTG CCCCACATGT GGCCCATACA	19620
TATACAGCCA CCCAATTAGA CAAGATGGAT GAAGCAAAGA AGTGCAGACC GACAGGAGCC	19680
GGATGTAGAT CGCTCCTGAG AGACACAGCC AGAATACAGC AAATACAGAG GCGAATGCCA	19740
GCAGCAAACC ACTGAACTGA GAATAGGACC CCCGTTGAAG GAATCAGAGA AAGAACTGGA	19800
AGATCTTGAA GGGGCTCGAG ACCCATATG TACAACAATG CTAAGCAACC AGAGCTTCCA	19860
GGGACTAAGC CACTACCTAA AGACTATACA TGGACTGACC CTGGACTCTG ACCTCATAGG	19920
TAGCAATGAA TATCCTAGTA AGAGCACCAG TGGAAGGAGA AGCCCTGGGT CCTGCTAAGA	19980
CTGAACCCCC AGTGAAC TAGTGGTGGG GGAGGGCGGC AATGGGGGGA GGGTTGGGAG	20040
GGGAACACCA TAAGGAAGGG GAGGGGGGAG GGGGATGTTT GCCCGGATAC CGAAAGGGAA	20100
TAACATCGAA ATGTATATAA GAATACTCAA GTTAATAAAA AAAAAAAAAA AAAAAGAGAT	20160

CACTGCTCTT	GCAGAGGCC	CCAGTTCTGT	TCCCCACAAC	CTCTTAGGAT	GACTCACAAC	20220
CACCTGTAAC	CCCATTTTCAG	GGGATCTGAT	GCCCTCTTCT	GGTCCCCATG	GGCACTGCAC	20280
TCATTTACAA	ATACTTTTCAC	ACAGAGACAC	ATGCACAGAA	ATGGAAATTT	AACAGAATAA	20340
ACATCGGAAC	ATTAAAAACA	AAACAAAACA	AAACAAAAAA	CAAAAAAAAC	CCCATAGGAC	20400
TGGAGAGATG	ACTCAGTGGT	TAAGAGCACT	GACTGCTCTT	CCAGAGGTCC	TGAGTTTAAA	20460
TCCCAGCAAC	TACATGGTGG	CTCACAACCA	TCTGTAATGG	TCTCTTCTGG	TGTGCCTGAA	20520
GACAGTGACA	GTGTACCCAC	ATACATGAAA	TAAATAAATC	TTTAAAAAAA	AAAAGCCCAG	20580
AAAGTGATGA	ACTCTATTAC	CACCAAAAAG	AAAAAAAAGA	AAAAGAAAAA	CTCAAATCAA	20640
TCTTGAAGTC	TCTTTTCGCAT	ATCTCTTTGG	CCTCCACCCT	GTCTGTGGAT	CCCACATGTG	20700
GGGGCGGTGG	GGCATCTGTG	TTCATTTGCT	GAGTGTGAGA	GCCACATAAA	GTGCTGGTTT	20760
ACGTGTTTAC	TTGTTTTCTA	GATGGGATGG	AGCCCAGGAC	CTTAACCTTG	TGGGGCAAGA	20820
CTTGCTCTCC	TGAGCTCTAC	CCAAGCAGTC	TGGATTGCGG	GTTTCCTGTT	TGTCTGTGAG	20880
CTCTCTGCTT	TGTGGTCATT	TGTGCCCCACT	GGCTCTTAGA	TCCCATGACT	TCCCAGAGAA	20940
CGCTGTCCTG	CAGAGGCAGA	CACAGGGCCC	CTGAGCTCAG	GCCCGGCCCT	GGAGACAGAA	21000
ACGGAGAGGC	CAGTTGATTC	TTGATATTTT	CCGTTGTGGC	TTCCTTGGGG	CCTGTGTGAG	21060
AATTCAGCTC	TGTAGAAACC	TTATGGTTCT	GCAACTACCC	TCCCCTGGCC	AAGACCCTTT	21120
CTTCTAGCCC	GGGATTACCC	CCTCAACCTC	TGAGGTCGCC	GCCAAGGTCT	CCTTCCAGAT	21180
ATGGAAGTGA	CTGGATGAGT	CCCTTGCGGC	CTCGCCTGCC	TTCCCATCAC	CCAGGCCCTT	21240
GTTTGGCTTT	GCCCTTTCCC	ACAGAAAGTCC	ACCATTGCTG	TTTGGACTTC	CAAATGGTGC	21300
TCCTAAGTCT	GTCTGCCGCA	GGCCTTACCC	CAGTCGGGAG	TGGGAAACGG	GCCTAACTGG	21360
AGATGACAAC	CTGTAGAGAC	CCCTCGGTCC	TCCTAGCAGC	CTGCTGGGCT	GTTCTCCCTC	21420
TGAATTGCCA	ATGTCCATGG	CGTTCCCGGT	GCCTTTCCTC	CCTCCCGTTT	CTGACAATTA	21480
GACGCCAGTC	AAGTTTGAAA	AGGAAATCTG	CTTTATTTAT	TTATTTATGT	TTTTTTTTTA	21540
ATTTTTTTCT	GGTAGTGGCC	ATGGGGAACG	AAGGAAGCGC	CCTAAAGGTA	TCATCACAAA	21600
GCAGGGCTCA	GCGGCCGGTC	TCAGTGCTGG	GAGAAGGCGG	CCTCAGGGTC	GCAGGCGGGG	21660
TCCTCGTGGC	AGCCGTCTGC	ACAGAGAGAC	GCCGAGTCAG	GGCCGCCCTG	GCCCAGCCCG	21720
CCTGGTCCTG	GAGCCCCGGT	CCCGTCTCGA	CCCCTGCCCC	ACTCACCCTG	GCTGCAGCAG	21780
ATGCCGGCGG	CGGCGCAGCG	GCCCCCGCTC	CCGCAGGGCT	TCTGGCCGGA	CTGGCAGGGC	21840
GACGGCAGGT	AGTTCTCCTC	TTGGCAGCGC	AGCGCCTCGG	CCGTGCCCAC	GAAGCAGCCC	21900
AGCTCGTCCC	CGCAGCAGAT	GCTGGGCCCC	AAGCAGCGGC	CTTTGCCCCC	GGGGCCGCAG	21960
GGGAGACACT	GGTGGGAGGG	AAGGGATGAG	CCGGGGGCGG	GAGGGGAGCG	GCCGGGGAGG	22020

GAGACCCTGT	GGGGCGGGG	GCTGAGCCGG	GCGGGCGAGG	GCGGCCGGAG	GAGCGCGGGA	22080
GGTGGCGGGT	CTCCCTGGCT	CTCTCTTTGG	GCTCAAAAGC	GGTCGAAGGA	GGGCAGTCAA	22140
AAGCTCCTCC	GCTCCCTCGA	TCCCCAGGCT	AGGTGGGGCC	GGTACGCGGT	CAGCGCGGGA	22200
AAGGGGGCGG	CGGGGGCGAC	CCTGTGGCAG	CGGGCCGGGC	AGCCCGGAGA	GCCACGGGTC	22260
GAGGGCGGGG	CTCTCACCGT	GCGCACGTCG	AGGTCCAGCA	CCGCGCGTTT	GCCGCCAGG	22320
GGGCAGTTCT	GAATGTAGCA	GGCGGAGGTC	AACGCCAGGA	GGCCGAGCAG	GCAGCAGGCG	22380
AGGCTGGAAC	CTGCCATGGC	GTTGGTGTTT	AGTCCGAGAT	CGGTGACCG	ATCCACCGTC	22440
GGTGATGGTT	TCTCCAGCCC	AGACCGACCT	TTTTATGCCT	TGTCCACTGC	CATGGTGGGG	22500
CCCAGTCTAA	GAGGGTGACT	GATGACTGG	TCACAGCCAG	GTCTCTTGGG	TCAAACGTGT	22560
CCACACTGTT	TAGAAGCAGG	CCCTTCATTT	GCAGGGTCTG	GGCTGGGGTC	AAGGTCACCG	22620
CCTCAGCTAA	TGACCTGAGC	TCAAAGGGA	CACAGCCTAG	AAGGGGAGGC	CTAAGCTACA	22680
AGAGGATAAA	GAGACTTGGA	GGGGGTAGAG	GTGCAGCCTA	GCCAAGAGCT	GTTTTTTCAT	22740
AGAAATCCAA	TACCTCAGAA	TGAGGTTGGA	TAGCGCAAGT	GGGTGAGGAA	GCCCTTACGT	22800
GGATCTAAAG	CTTAGATGGG	GAAAAGGATC	TTGTTCAATC	TCTGAGTGCA	GCTCAGCCCT	22860
TCTTCTAACT	AGCCCGTAAA	ACAAAATATC	AGTAGAAATC	AAACCCAAAA	ACACAACAAA	22920
CAGACCAAAA	TAAAGTAAAA	AGAAAGAAAA	ATCACAATAA	AAGGAAAAAT	CACACTTGCA	22980
CTTACAACCT	TGTATTAGGG	CTGGAGAGAT	GGCTCAGTGG	TTAGGAGCAC	TGACTGCTCT	23040
ACCAAAGGCC	CTGAGTTCAA	ATCCCAGCAA	CCATATGGTG	GCTCACAACC	ATCTGTAATG	23100
GGATCTGATG	CCCTCTTCTG	GTGTGTCTGA	AGAGAGCTAC	AATGTGCTTA	TATATAATAA	23160
ATAACTTTAA	AACTCTTTAA	AACTCTGTAT	TAGAACTTGC	TATGAGGACC	AGGCTGGCCT	23220
TGAACTCACA	GTGATCTATT	TGCTTCTGCC	TCCCAAATGC	TAGGTACCTA	CACTCCCGTT	23280
TGAGAAAACA	CAGGCCATCA	GCTGCTTGAG	CGTGGCCAAC	AGGCGGCCTC	AGCTACAGAG	23340
AGCCATTTGT	CCTAAGGCCA	TACCCTTCCT	GGTGGCCACA	TGTAATGGTG	GCCCATTTTA	23400
GTACATACAA	CTAGGCATCT	CGTGTTGCAT	TTCAGGGTTG	GGCTGCAGGC	CTGCATAGGT	23460
CTGCATGGGA	AGAATGCTAC	ATGCAGCTCA	GTAGCAGACT	GCCTGCCTAG	TGTGTGAGAG	23520
ACCTTGGGTC	CAGTCCCCAG	CATGGTGGTA	GCAAGACATT	TTGGGAACAG	TTTTTGCTTT	23580
AAATTTTAAC	TTTTATTTGT	GTGTATGTTT	GTGTACACAG	GTTTCCTTGG	CAACCAGAAG	23640
AGGATGTTGT	ATCCCCAGGA	CGTGTCATTA	AAGGGGATCA	TGAGTAGGCC	TATGTGGGTG	23700
CTGGGAACAA	AATTCAGAAT	TCTGCAAGAG	CAGTGTGCAG	ACTTAACCAT	TAAGCCATCT	23760
CCCCAGCTCC	TTGATTTTGC	ATTTGAATAC	AGTTTAAAAT	GAAATGCACA	TTAAGCCACG	23820

GTGACAGTGA	TGCAAACCTTT	TAATCCCAGA	ACTCAGGAGG	CAGAGGCAGG	AGAATGTCTG	23880
TGAGTTCCAG	GCCAGCCTGG	TCTACAGACC	TAGTTCCAGG	CCAGCCTGGG	CTACACAAAA	23940
AAACAAAAGC	AAAACCAAAA	CAAATAAAG	ACACAGACAA	ACCATGGCAG	GAAGACATGG	24000
GAGCCTCAAC	CTCTTCATTT	GACGGCTGAG	AAATCGAAAA	CAGATGACCA	GGAGAGACCA	24060
AGGTCTCACT	GCTGCCTTCA	AGGCTTGCCC	TCAGTGACTG	GAAGATGTTC	CACTGGGCCG	24120
CATCTTAATA	TCTTACCATC	TCAGGGCTGG	AGAACTGGCT	GAGTGGTTGA	GTTGCTCTTG	24180
CAGAAATCCT	AGGTTTGATT	CCGAGGACCC	ACAGGGTGGC	TGAGATCACT	TATCCCAGTT	24240
CCAGTGGAGC	CAGTACCCAA	ATAGTGCATT	ACACACTTGC	AGGCAGAACG	TTCAGACAÇA	24300
TAAAATAAAA	TAAATAGACC	TAAAAACATT	TAAAAGAAAG	GAGAAGCATT	ATCCAGAGTC	24360
GTTTTATTTT	GTTTTGAGAT	AGACTCTTAG	TTGACCTGGG	ACTGTCTGTG	TAGACTAGGC	24420
TGGGCTTGAA	CTCACAGCGA	TCCCCCTGCC	TCCCAAAGTG	CTGGGTGTAC	CACCGTGCCA	24480
GGTACCTAGG	CCCCTGTTTA	AGAAGACACT	TGCCATCAGT	GGCTGGGTGT	GGTCTTAGCT	24540
GCAGAAAGCC	ACCTGGCCCT	TCCCAGGTGT	CCACATATAA	TGTTTGGTCC	ACTTTGGTAC	24600
GAATGCTGGG	CACCCCAACT	GCATGTCAGC	TTTGGGCTTT	GGGTTAGCTG	AGGTCTGCAT	24660
ACTGGTTCTA	GTTGCCACCC	CCTTCTCTTC	CATAGAGGTG	GGGCCTAAGC	CCGTGTTCTA	24720
AACTCCATCT	CAGGCTCTCT	TAAGAAGTGA	CCTGCGACAT	CCAGGAAGAA	GTAACAGCCA	24780
GTGCCCCCGA	GACCCACTCA	CTACATGCAG	TCTCAGCCCC	TAGAGAGGAT	GGAAAAGCCT	24840
CCGGTCTCCT	TGTTCTTATG	ATCAGCCTTC	TCCTCAAGGA	GCTGGGGCCA	GTGGGGCAAA	24900
GCACATTCTC	TTCTGACCCT	GAATCACAGA	TCCTGAGTCA	CTGGTGCAAA	CTATCAAGCG	24960
CTAAGTTGGT	GGTGAGGTTG	ACCTGTACTA	CAAATCACTT	CATTTCTCAC	CCAGACTAGC	25020
TTATTGGCAT	TCCAGGCATA	GAAAGCCAAG	AGCTTGACCC	CCACTATAGC	CCCAGAGAGA	25080
CAGCCCACAT	AGTCTGTGGG	CATAGTGATC	TCATCTTAGG	TAATCCATGC	ACATAAATTA	25140
GCATGTCTTG	ATAATACATA	CCTAATGCTC	CTGTTAGGCC	AGCATGCCTA	ACATGCTCAC	25200
CAACCCAATC	TGTGTTTGGG	AAAGGCCAAT	ATTCCGCAAG	GCAGAATGCT	AGTCCTTCAG	25260
GAATGGGGCT	GCAGCTGGAC	TGGGGAGAAC	ACACTGAGGT	TATAAGAGGA	CCATTGAGGC	25320
CTAATAGCCA	AGGTAGAGTA	GGCGGAGCCT	TGGGTTACAG	TGTTCAGCAC	CAGGAGGAAA	25380
GAGTCACTAT	CACCATGGGG	TTCATCTGTC	ACTGGAGGAA	GCAGAATATG	AACTAAGAGG	25440
CATATTATGT	TGGGTTACGA	CTTTAGTTAA	GATCTGAGTG	TATCCCATGT	GATACATTGT	25500
CAGTCCTTAG	GAAGATGTCT	TGGGAGATGG	TGAGATCTTT	AAGATAGAGA	CCCAGTGTCA	25560
GGTTCTTTAT	GTCTCTGACA	GCATGCCCAT	GAAGGAAGTG	GTCTCTCCTG	GATCTCTTTT	25620
TCAGTTTTGC	AGGCATGGGA	TGAAGGGGTG	TATCCTTCCG	TGTGCTTCTG	CCATGATGTG	25680

TACTTCAACA	TAGACCTGTA	AGGAACAGTG	GCTACAGATT	GTGGACTGAA	GTCTCTGAGA	25740
CTGTGAGTCC	AAATAACCCT	TTCTTTCTAG	GCATGGCGGC	ACACACCTGT	AATCTCAGCA	25800
TGCTGGAAAT	GTGCAGCAGG	ATCAGGAGTT	AAAGACCAGT	CTCAGATAAA	TGACAGTTCA	25860
AAGCCATCAA	GGGGATAATG	AGATACTTCC	TCAAAAACCA	TCAAATTAAA	ACTTTTGTTT	25920
TTATACATTA	CAACTTGTCA	GGGGTTTTGC	TATAGTAATT	AAAAGTCACC	ACAGGAAACA	25980
AAGGCACGTA	AACATAGCAA	CATGTGCTAT	GTTTAAGGCA	ACATGTGCTA	GGAAGGTAGA	26040
TATCACCATG	CTGGGTGCTT	AGACCAGGGC	TATGTGAGG	TCCCGGAGGA	GAGCTGAGGA	26100
AGCCCTGGGT	GAATGTATAA	TGTATCACGG	GCCTCAGACC	TGTGAGATCT	GGCAAAGCTT	26160
CCCCCTGCAC	GCTGTGGGTG	AGGTGAATGG	GGATTCGGCA	GAGCCTTTGT	CTGGTCTGAG	26220
TGCAAATGCT	GACGGTATGT	TCTAGTGGAG	GTGTTTACAA	AGGACGGGCC	AGTGTGCGCT	26280
TTAGCCATAG	AAGTGGTGGC	TCCCTGATGA	ATGTCCACAA	CCTGGGATTG	CTGCCCACAA	26340
GATCAGCCAG	GCCCTCTCCT	GCGCTGTGCA	GAGTGAACAC	ACGGAGGTTT	TGGGCTGCTC	26400
CAGTGGCTGC	TACCATTCTG	CCAGAGAGTG	CACAGGCCAC	CTGACCCAG	CCTTTCTGTC	26460
CATGTGTCTG	TCCTTTCTTC	ACTCTCTCAC	CACCCTTGTT	AGGGTCCCAG	ATCCAAGTTA	26520
TGTAGGGGGT	GGATTAGGAA	ATGCTATGGG	ATGAGAGGCA	GTGTTGGTTG	TCATTCTCCT	26580
TAGGGTAACC	TGTGAGTATC	AAGGAAAGAA	AGTGTAACG	CAGAAGGCTC	ACCGTGCTGC	26640
TGCTATGTAC	AAGTGAGCAC	AAATGTAACC	TCTGGAAATA	CCCATTTATC	ATGTCTGTTT	26700
TGGGGGCAGA	GCCCAGGCAG	GCGTTTCTAC	TCATGGTCCT	AGGAGCAGCC	TCTCCTCATC	26760
TGGTATGCAG	CCCTTCCTTA	TCCGAGACGG	AGCCTGGTGC	CGGGACACAG	GTCATTTCCC	26820
TGCAGTTGTA	TATTATTTGG	GCAGCTCACT	TCTTTAAAT	ATTTTGA	AAATTATGTG	26880
TATGAGCCTG	CATGTATGTC	TGTGCAAAAT	GTCCACAGAG	GCCAGAAGAA	GGTGTGAGAC	26940
CCCCTGGAAC	TGGGAGTTCC	GGGTGGTCGT	GTTTGGCATA	TGGGGCCTGG	AAAATGAACC	27000
CTGGTCTCCT	AGAAGAGCAA	CCAGTGCGCT	CAGCTGCTGA	GCACCTCTCC	AACTCCTGCT	27060
TCTCTGGACT	GGGAGACAAA	GGAAAAGTGA	GAGACTGATT	CTGTTCTGTC	AAGTCTCTGA	27120
GCATAGGGAA	GACCTAGGTT	CATTCTATGT	CATCTGTCTG	TCTGTCTGTC	TGTCTGTCTA	27180
TCTATCTATC	TATCTATCTA	TCTATCTATC	TATCTATCTA	TCTGAGACAG	GATTTCACTA	27240
TGTTAGCCTT	GGCTGTCCTG	GAACCTATG	TAGACAAGGC	AGGTCTTAAA	CTCGCAGAAG	27300
ATCCTGGTGG	TCTCTCCCCA	CTTTGCCTGA	TTAGGCTCAC	TTTTAAAGGG	AATGAAATGG	27360
GCTGGGTGTG	GCAGTACACA	CCTTGCTCTC	AGCACTCCGA	GGCACAGAAA	GGCAGATCTC	27420
TGAGTTTGGG	GCCAGCCTGG	TCTATGCAGT	GAGCTATAGG	CAAGCCAGGG	CTACATGGTA	27480

GGACCTTGTC	TTAAAAAGAG	CCCCAAACAA	ATAGCTCACT	TGCCCAGGTG	AGGTCCACCA	27540
GCATCTCTAC	ATTTTGACCG	GAAGCTAAGA	GGAATCTTTA	TTACATCACG	CCTGCCACAG	27600
TCTCCATCTT	TGTTGCAGCT	GGAGTGCTCC	CACAGGGCTT	CCACTGCACG	CACTGCACCC	27660
GAAGGGGCTT	CCACTTCACG	CACTTCACCC	GAAGGGGCTT	CCACTTCACG	CACTTCACCC	27720
GAAGGGGCTT	ACACTTGATT	CACTTGACCC	GAAGGGGCTG	ACACTGCTTG	CACTGCACCT	27780
TAAGGGGCTG	ACACTGACCC	AATGGCACCC	GAAGGGGCTG	ACACTGACCG	CACTGCACCG	27840
AAGGGGCTGA	CATTGCACAC	GCTGCACCCA	AAGGGGCTGA	CACTTGCTGC	ACTGCACCCC	27900
AAGGGGCTGA	CACTTGCACG	CACTGCACCT	ACCAAGGGTG	ACACTGCACC	TGCTGCACCC	27960
AAGGGGGCTG	ACACTGCATG	CACTGCACCT	ACGGGGGCTG	ATACTGCACC	CACTGCACCC	28020
AGGGGGGCTG	ACACTGCACC	CACTGCACCC	AGGGGGGCTG	ACATTGCACA	TGCTGCACCC	28080
AAAGGGGCTG	ACACAGCACC	CACTGCACCC	GAGGGAGCTG	ACACTGCACG	CACTGCACCT	28140
ACCGGGGCTG	ACACTGCACC	GCTTGTAATG	TACATTACTG	TTTTTTTTTT	TTCTTTTCTT	28200
TTTTTCAGAG	CTGAGGACCG	AACCCAGGGC	CTTGCCCTTG	CTAGGCAAGT	GCTCTACCGC	28260
TGAGCTAAAT	CCCCTACCCC	TACATTACTG	TTTAGAAACA	AATTTATGGT	CCTTCTCACA	28320
TGCTGCAGGA	GATTACACAA	AGTTGGGGGT	TATCAAGAAT	GTGGATCACG	GTGGATCATT	28380
TTAGCACTGT	CCCCCCCACA	GAAAGGGTCA	TTTCTAGACA	GAAGAAAATA	GTTTATATGG	28440
AACACTTCTG	GGCTGGGCAG	TGGTAGCACA	TGCCTTAAAT	CCCAGCGCTT	GGGAGGCAGA	28500
AGCAGGCGGA	AGCACGCGGA	TGCACGCGGA	CGCACGCATA	TCTGTGAGGT	TTAGGCCAAC	28560
TCGGTCTATG	CAGCAGCTTC	CAAGACAGCC	AAGGCTGTAT	GGAGACCCTG	TCTCGGGGTT	28620
GGTGGGGAAT	CTCTTCACCG	TCTTGGTCAC	TTCTTTATGT	GTGAGACACA	TAGACGTTTT	28680
TCTTCTGAAT	ATTTTATTGC	TGCTTGTGGC	ATTCACAAC	TAGGGAAAAA	TTGTTAAATG	28740
CTGCATTCCC	AGCACTTGAG	CCAGTGAAGT	TCAGGCCTCC	GCTCGTCTTG	TAATGGTATT	28800
TGCACAGGGG	ATGCCTTGGC	TGAGTGAGTT	CTTCCAGAAA	ACTCCTGGGC	CCTTAACACC	28860
TATTTCCAGC	ATTTGGAAAT	CCGAGGCAGG	AGGATTGACA	TGAGTTGCAG	ACATAGTCAG	28920
CTAGAAGTGC	AGCATTAAAT	CCTATCTTAA	AATAATTATT	AGAATAATTT	AGGGGGAAAA	28980
GCCTCTAATA	GAGATGGGAG	AGTGTGCGCA	TGACTGCCCT	ACTGTGTGCT	TCTAGAAATC	29040
AATATGAATG	GGCCAGAACT	AGAGAAAAGG	CTGTGAGAGG	CTGTACCCTA	CTGTGTGCAA	29100
CCCACTTCCC	TCCTACTATG	TGGGTGCTGG	GCATGACACG	AGGTTATCAG	GCTGGGTGAC	29160
AAGCACCTT	ACCTGTGGGC	AGTCTTGCTG	GTCCAACCTA	TTTGCAATTT	AATCCCAGCT	29220
ACTTCAAACC	CCATGGGTGC	ATATTTACCC	ACTTTTGGTT	TTGGAAACAG	GATCTTAAAA	29280
TAAACAGGTC	TCACTCTGTA	ACCCATGCTG	GCCTGAATTC	AGCATCTTCA	GCCTCAGTCT	29340

CCCAAGCGCT	ACGATTTCCT	ATGTGCCATA	TGTCACAATA	CATGCACTTC	AGTTTTGTCA	29400
AAAGAAGTGA	ACCAGGAATA	ACTGGTACCT	ACCTATAAGA	CTGCTGTGAT	GAAGGAGGAC	29460
ATTGTGTAAA	ACGAAACTCA	GGATATAGTA	AGTGCTCAAC	ACGTGTTAGA	CATGTTGGTC	29520
TCCATGAGGG	CACAAACCCA	GGGCCTCATG	CATGCCAAGA	ATTGGCCCTA	TCCTGAGCT	29580
ATACAATTAG	TCCCTATGAC	CTACTGTGAC	CTCAGACGCA	CACCATGGAT	CTGACATTGC	29640
ATCAAATCAG	AAATGAATTT	CTGAAAGACT	TGCTCATAGC	ATGCCCTCCC	ACACCCCCGT	29700
CCCAGCCCCC	CCTCTCACTG	GCAAGGACAT	CTCACTGTGG	TGGTGGCAGG	GCCTCTAAAA	29760
CATCATAGGA	TAGCTGAGCA	GCAGTGGCAC	ATGGCCTCTC	AGTCCCAGCA	CAGGGGAGGC	29820
AGTCAGCCTG	GTCTATAGTG	TAAGCTCCGG	GACAGCCAGG	GCTACATAGA	GAAACCCTGT	29880
CAAACCTACC	CTACTTAAAA	ACAGAAGTAG	AGCAGTAGTT	TGGATTACAC	TGCTTTGACA	29940
CTTGGTGGGT	AGCATGTGTG	CACCTGCCCA	GGAGCTATCT	GGATTCTCAA	ATGGAAGACA	30000
CAGACACAGA	CACAAACACA	AACACACACA	CACACACACA	CACACACACA	CACACACACA	30060
CACACACACA	CACACCAGTT	AACTTTTGAC	ACGCCATGAC	TAGCTCAAAG	GCTAGGGACT	30120
CCCAAACCTT	CCCCTGTCAG	CAAATGCTCC	CCTCTGGTAC	TCCTGAGACT	AAGCTAAGCC	30180
TTCCCCTGCT	GTCCCAGGCC	CAACGGAGGA	AGTGAGCATG	GTCACCTACC	TGATTCTTTT	30240
TTTTCTTTTT	TTCGGAGCTG	GGGACCAAAC	CCAGGGCTTG	CGCTTGCTAG	GCAAGCGTTC	30300
TACCATGAGC	CAAATCCCCA	ACCCACATA	TTCTGATTCT	TACATGGCTG	ATTGGCTTTC	30360
TGTCCCTGCA	GTTCTTACAT	CCTGTCCTTC	TTCCCTGAAT	CATGAGGACC	CTCTCCTCTC	30420
TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCCCTT	CTCTCTGTGT	30480
CTCTGTGTCT	CTGTCTGTCT	GTCTGTCACA	CACACACACA	CACACACACA	CACACACACA	30540
CACACTAGCC	CATGCAAATC	TAAGGGCCCC	TTCCCGTCTC	CCTTTGCCTG	ACCATTGGCT	30600
CCTGGCATCT	TTATTGATCA	ATCAAAAACC	AATTGGGGAT	AAGGACCTAC	AGTGTTTGGA	30660
CATGAAGATT	CCTAATTTGG	GGGCTGCATT	AATTCAAAAC	ATTGGAACCA	ATTCCCAACA	30720
ACAACAACAA	CAACAAATAA	AGCAAAGAAA	AAAGTTTACA	ACGCTGCCTT	CATTATTTGA	30780
GAAACAAATG	TAAGGAAAAC	CATCAGGTAT	CTGGACTTTT	AAACGGCCGG	GATTTAGAGA	30840
CTCTGGGATG	TTTTCTGTGT	TGGGGATTGA	AGCCAGGGCC	CTGGGACATG	GTAAGGAAGC	30900
ACTGTACCAT	GAAACTACAC	CCCAGAGTCT	GATAAGGCTA	CTGAATGACA	ATTAAAGATT	30960
CATAATTGCT	GAAATTCTGG	AAAACCTCTA	GCTACCAATT	TTGTATATGC	TCAACTTGGT	31020
TTCCTGAAAA	CATCTGAGGT	TCTTGACAGT	AACTTTTCCT	CAGAGCAAGT	ACAACATAAT	31080
TCTGACTTTG	TGACAATAAA	GATTGTCAGG	AAAGGCTTTG	TGAAAATGTT	CAGTCCCCAG	31140

GAGACGTGCC	CTCCTGCAGC	CTGTGAATGG	CGGCCAGGTC	ACAAGTCAGC	AGATGCAGTG	31200
GAACGGAGTG	TGGTACTTCT	GTGAGACACT	GCAGGACTGG	ATGGATGGCT	TAGTAGTTAA	31260
GAACATGGGC	TGCTCTCCCA	GAGGACCTGG	TTTCAATTCC	AAGCCTTGGG	CCCTGCAAGA	31320
ACCTTATATA	GA CTGGCTTC	AAGTTCTCCA	TGTAGCTGAA	TATGACATTA	AACTCCAGAT	31380
ACTCCTGAGT	CCTAGGTTTA	CAGCTGTGTA	CAGCTATGTT	TCTTCCCTGA	CGACCCCGCA	31440
GCCCCCATTT	TGAGATAGGG	ATTTAGGTAG	CCCAGGCTGG	CCTCACACTG	ACTAAGTGAG	31500
ACTGGCTTTA	AACTCCTCAT	CCTTTAAGGT	ACCACCATGA	ATTTGCTGTA	TAGCTCTGGC	31560
TGGCCTTATA	TAATGTAGAC	TAGACTGGCC	TTTAACTTTA	AAATTGTGCA	CTTTATTTTT	31620
TTTTAAATTA	TGTATTTTAT	GTATATGAGT	ATACTGTAGC	TGTCTTCAGA	CACAGGGCAC	31680
CAGACCTCAT	TACAGATGGT	TGTGAGCCAC	CATGTGGTTG	CTGGGAGCTC	AACTCAGGAC	31740
GTCTGGAAGA	GCAGTCAGTG	TTCTTAACCT	CTGAGCCATC	TCTCCAGCTC	TCTGCTTAAT	31800
AAGTGCTGGG	ATGACTAGCA	TGTGTCACCC	TCCTGGCCAC	TTCTGGTGTC	TCCTTTCCAG	31860
GCTTTTAAAA	ATTATCTGTT	GGCATGTCCA	CACAGGGTTA	TATGCATATG	AACGCAGGTG	31920
CCTGTGGGCT	GTCCTGTCCT	GGA ACTGGAC	TTACAGATGG	CTGTGAGCCA	CTTGATGTGG	31980
GTGCCTGGAA	ACTAACTGGG	GGTCTGAAAA	AGCGGGAAGA	ACTCACATGA	CTGTGGAGTC	32040
TGCTACCCCT	TTTATTATAA	AAGAAAAGAA	GATATTTTAA	CAGCACGTAT	GAGACACAAG	32100
TGAAAGCTGT	GGCCATGGTC	TTCAGGGATG	GTTAGGTCCT	GCAAACTGA	AGGAGGTGGG	32160
CTCTGGGTGT	TGGTCACATG	G TAGATTGAT	AGGCCCTGGG	TTCAATCCCC	ACCTCTGCAT	32220
AAAGCAGGCA	TGGTGGTTCA	CTGCCTGCTT	TTAGAGGAAG	AGGCAAGAGG	ATTGGTAGAA	32280
TCTCAAGGTC	ATTTTCAGCT	ACATAGCACC	AGTCAAATCT	TTGAGTCCAA	GACCAGCCTG	32340
ATCTGTGTAC	TGAGTTCCAG	GGCAGCTACA	TAGTTGAGAC	CCTACTTAAA	ATTTCAAACA	32400
ACAAAACCCA	CAAGGTTTAA	AAACTCTATC	ACTTTTAGTT	ATGTTTGTGT	GTAAGTGTTT	32460
GTGCCCACGG	AGGTTAGAGC	ACTGTATCCC	CCGGAGTGGT	GAGCGGGCTG	GCATGAGTGC	32520
TGAGGGCTGA	ATTCAGCATC	TGTGTTCAAC	ATATTTGTTA	AAGCACACGA	AGAGGAAATG	32580
GCCAGTGTCA	ACAGGAGCCC	AGCCAGGCTT	GGGGTGGGAA	AATGCTTTGA	CTTCTATCTG	32640
GCAAGAAAAA	AACAATTCCA	AGTTTGATCC	TTGCCAGACT	CTTTGGCCTT	TACCAGGCTT	32700
CCTCACAGAG	TCTGCTGTAA	CTGTTTCTGC	AAATTCGCAG	AGGAACCTGA	GATCTCAGGG	32760
CACGTTGGAT	ACCCACGTGC	TGGAGAAACT	GAA CAATGAC	TTAGGTTTC	ATCGTGCCTG	32820
GATGAAACAT	GAAAATACCC	CACACCGCTG	AGCTGACAAA	TGTGCCTCTC	TCTCTGTAGC	32880
CCTTCAGTCA	GCTGAGCAGT	TTGCCCTCGC	TCGGCTGCAG	TACCAGCACA	GGCACCCAG	32940
TCTCCCCGAT	GAGCGGTCTC	ATGAGGATTC	AGGTTGGTCT	CGAACTCCCT	ATGTAGCTGA	33000

AAATGACCTT	GAGATTCTAC	CAATCCTCTT	GCCTCTTCCT	CTAAAAGCAG	GCAGTGAACC	33060
ACCATGCCTG	CTTTATGCAG	AGGTGGGGAT	TGAACCCAGG	GCCTATCAAT	CTACCAAGTG	33120
AACAACACCC	AGAGCCCACC	TCCTTCAGTT	TTGTAGGACC	TAACCTAGCC	CTGAAGGCCA	33180
TGGCCATGGC	TTTCCCCTCA	GCACCCACTT	ATCATGAAGG	GGCAAGGGTC	CAGTTTCTTG	33240
GTTAAGTATC	TACGCTTGTG	ACTAGGGAGA	TACATCCTGG	GCAGGAGTGA	AGGGTTACCC	33300
ATTCAGCAGC	AGAGTTCCCTA	GGTTTACTGT	GACAACAAAG	ATCTAGGAAT	GGCCTAGGTT	33360
GTCCTGACAT	GATCCCATTA	GCCTACCTCA	GATATCTGAA	TGCAGGGGCT	CACTGTGTGT	33420
CCCAGTCAGG	GACAGTATTT	ACTACCCTAA	AGTGGGTTAC	AGCTCTCGGG	GGGGGGGGGC	33480
TGCGTGCAGG	ACGACACCTG	CACCTTCACA	CTTGCTTCTT	CAATGGAGTA	AGAGGCTGCT	33540
AACATCCCCA	AGGTTTCCAT	TTCAGCTAGG	ATGAGAGTCT	GGAGTTCATG	TCCCTGGTAT	33600
TCAAGTATAT	GACACTGAAG	AGCAAAGAGG	CAGAGAGCTC	ATCCACTAAC	AGGCATGCAC	33660
TGCACTCATG	AACATCTGTG	TCTGATCCCC	GGTACACATC	AAAGCCATGT	GCACCGAATT	33720
CCAGCGCCAG	GCAAGCATAC	GCAGATGCAT	TCTCTGAGGC	TAGCTGGCTG	GGCAGTCTAA	33780
ATACTGAGTG	CCAGGTTCCA	GTGAAAGGCC	TCGTCTCAAA	AACCAGAACA	AACCAAATCA	33840
AACCAAACGT	AAAGCAGACC	AAGGTGAGAG	AGTCTTGAAA	TGACACCCAA	GGGTGCTGTC	33900
TGGCCACTGC	TTACACACAC	TGAAGGACAG	CAACTGACCG	CAAGAAGCGG	GTTTAGAGTG	33960
GAGTCTACTG	TCTGCTGGGT	AGTCCAATGA	CGCTGTGTCA	GGGCAGGGTC	CGGTTACAGA	34020
AATCACTGAC	GGGGAAGCCT	TCCCAGGAGA	AACGGGGCAC	CCTTTTTGCT	TTCTGGACCT	34080
TGGACACACC	TGACCCTACC	CAGCAAAGCC	CAGGATTGAG	CAAAGCAGAT	AACTAACTCC	34140
TGGCTCAGTT	AGGTGAACTG	GCTTTTGGCT	AATAACCTTA	AGACCCAAAT	AACTGGGACA	34200
AATAAACTTA	TTCTACAACA	AGAAAAAGCA	AGCCACATAA	CAAAAGGCTT	TGCTTTCCAA	34260
TAGTTTATTT	ATAAAAGCAG	GAAACATTGG	GCTCACACTA	TTCCAAGAAA	CTCAGGAGAC	34320
AGCTCTTGGC	TTCTAGAGGG	GACAGTCCAG	TCTGATCTTC	TGTGTGATAG	GAACCTCCCA	34380
GAGTTTAATG	CTGTCCGGAG	TTTGTATGTG	GTGTCAGAAC	AGGTACTATA	CCTGGTGCCA	34440
AATCTAGAAT	GAATGGGGGC	TGCTCTGGAC	AAGGCTGTGC	CACCCTCCTG	GGATGCACCA	34500
ATTTCTACCC	CAATATCCAG	TCCCTAGGCA	AGCACTGTCC	TTATTTGGGC	CTGAGAGGCC	34560
TCTGTTCTCA	GTTCTTGACC	ACCTGCCCTT	CTTTGGTGAC	CGCCAGTTA	TAACCTTCG	34620
GAGAGTCCAA	GGCCTCTTCT	ATCCGTGCCT	CCAGGTTCTC	CCGAGTGATG	AAGTTTTTGG	34680
CCTCCTCCTA	CGGGAAGAGA	AAGGTTAGCT	ACGTGGGATC	TCTCAGAGCT	GGCTGCCTGC	34740
ATAACTTACT	GTCTCTAGCC	CACCTTAAGA	AGGGTGTCAA	AAGAGCCCAG	GGCCCCTTTG	34800

AGTGTCCCAC CAATAGGCCA GCACAACTGT ACTAGGTGAC CTCAACCATT ATATCCCTCT	34860
TCCTCGGACA ATTACCCATC AGATCTCTCA GCCACAACCC AAGAACGAAT CGAATCGTGA	34920
CAACCTGATT GGGGGCGGGG GGAGGTGCGG GGTGGTGGAC TCTCTCAACC TTGTCACTCT	34980
CCCTTACACA ATAAAAACAG CTCCATTCTC ATTTATGGTT TCCCTGACCT TCAAGTTTCA	35040
CACGGCAATA CTAGCCTGTA ACCTTTCATT GACCATTGTC TGGGAGGCAC TCTGTGTGTG	35100
CAGAGCAGAG AGAGAAGCCT CCTGAAGACT GGTCACTTGA CACCCTGTGC TGCTCTCCCC	35160
AGAGCCTCCC TCACCGGAAT CTCCAATACC CACATTCTC ACGACCTCGG CCCACCTGCA	35220
GTTTGAGAAC TTCTTGTTCT TTCAGTTGCA CCCAAGCCTG CTCCTCCTGG GCCCTCTGGG	35280
CCTGGACCTC AGCCTGCCGC AGCTCCTGGG CCTGTGCTTC GAGCTGCAAC CTAGCTATCC	35340
TGCGAGGAAA AAGGAGCCGA GAATGAGCGG GGCCGTGGGC CGCGCACGGG CAGAAGGAGC	35400
AGGGTTGGGG GCTGGCCCCC GGGCCGCGCG GCCAGGCGGT TCAAAGCCCG CAACATGACA	35460
GCTAGCGCCT GCGCCCTGTC CGGAAGTGGT CGAGGAGCAT CACGGGAATC TCCTGTCCCC	35520
GCCCCGCTAG GCGGAGCTCT TGCTGTCTGG CCCATCGCGT CCCTGGAGTC TCGGGTAGAG	35580
TGGGAGGGTC GAGGGGCACC TTGGAGACCA CAGTTCGCTG GGCTGTATTG CCCGCCAATG	35640
ACAGCTACAA CACCAGGCGC TCTTGCTTAC TTGTTACAA GCTTGGACAC CGCCCCAAAG	35700
CGTGTGGTTC TCTGGCCCAC CGGTGGCCTG TGTACCCTTC CAGGGCTGAT GCAAGAAATT	35760
CTGAGCTTCT AGCTACTTAA GTTAGGCTTA CACTTCCTCT TTGAGTAAGC GTTCTCCTGA	35820
TGCACTTTAA ACTTTAGGTT TGCAATCGAG CAACAAACCT TTTCCTTGCA TTGCACTGCC	35880
CTGTTGTGGT CTTGGGCAGC TCTGCAATTA ATTGCCAGAG TTCTCAGTAT CAGTGATCAC	35940
TTGTTTTCTA CTGGGTGGCC TGTGTGAGGT GACGCTTGTC CTCTTTCACC TGAGCCGGGG	36000
TCTTCCCACC TTGAGACCAT TTATGAAGCT TTTAAAGTAC TATCTTCCTC TACAATGAAG	36060
AGAAACAGGA GGCAGTGCAG GGCAGAGAGC ATCTGTCCCA AAGGTAGATG GCTCTCCTCT	36120
GGGAGAGTTG TTGGTAATCA GAAGCTGACG GTGAGGGCTA GGGAAGTAA GTCTCATCCT	36180
ACTCATTAGT GTTGTCAACA CCAGATTCTG CAGAGAATCT GCACTTAGGA GGGTCTGAGG	36240
AGCCTGGGCT GACTGCAATA CCTGATATTT CTCAGAAGAG AACAGAGTCT GCTTACATCC	36300
TTCTCCCAAC TTCCAGGCCC GTAGGAGGCC CAGCCAGCAC CCTCTTCCTC TCATCTCACC	36360
CTACCCTGGA GTGACACAGG GTTGCTCTGA ACATCAGGGA ACGTGGCACT CCCATCCTTT	36420
CTTGCAACAG GTCTCACTAT ATAGCTCCGG ATGGTCTGTC ACTTACAATG TATATCAGAC	36480
TCACAAGAGG TCCATCTGCC ATTGCCTCCT AAATGCTGGG GTTAAAGGCA CATAACACCA	36540
CACCTGTCCT AAACCTTTCT TCTTCGGGGT CATCCTAGAT AACCAGTATC TCATTTTACA	36600
TAACTTCAGT GTCTGGGCAA AGAGAATATT TCTATGGTGT GGGTCATTCC TAGAGGCTTC	36660

CTAACCTTGC TGGCTCTGAC GTTCTCTCGG CTGGTCAGGT CTACTCATCC TTCTTTCAGA	36720
GGGTTTCATA AGTTGTAAGA GATTTAGGCC TACGGTGGAT GAAAGATGTG GAGTCATTTT	36780
GAGTAGCTAA TGCTACAGAA CTAGAAGGCA GGTTCCTCTGC CCCCTTCTCT GACCTGTTGG	36840
GGAAGTGGAA GTAACCTTCC ATTTGTGACC TTCCCCACTA GGTGGCGAGA TAGAATTGTC	36900
AAAGCTGGGA AAGGAGGCTT TTCTGGGCAG TTCATGGGTA GAAGGACAGA CAGACAGAGT	36960
TGGAGGAATG GAAGCCTCCT CATTTACCAA GGGGTAAACT ATGGATGAGG TGACTTCAGG	37020
TGCCTGCAGG ACCCTATGCA GACGGTCCCA GGATTTAATG ATCAGGCCAT TCTATTTCTT	37080
CTGGTGTCAA ATCCAGTGAT ATCATTAAAA CAAAAACAAA AAAGCCCCAA TCAGGGTCTT	37140
ACTTGATGGC CTTATATTTT CAACAAAGCC CAGGCTGGCC TTGAACTGA AGCAATATCC	37200
CTGCATCTGT CTCCAGAGTG CTAAGATTGT GTGTGTCACC ATACCAAGGT ACAGTGATCT	37260
CTTGAAACAG GGAGGTGCAA GTCATTACTC AAACCCCTCC TCACAATGTT CTATGAGCAA	37320
ATCCGAAGTT GATGTTGGCT TTTAAAGTCA CCAGACAAGT GTCCTTCTGC TTAGATCTTC	37380
CTAGGAACTG AGGTTTGAAA CAAAAAGCAT AACATGGTTG GAGAGATGGC TCAGTAGTGA	37440
AATTCTGAAT GTGGTTCCCA GCATCCACAT TGGGCACCTC AGAATGGCCT ATAACTTCAA	37500
TTCTAGGGAC CAAGTAATCT CTTCTGGCTT ATGGGTGTCA CTCACATGCA TGTGTGCATA	37560
TGGTGCCTAA GTAAAAAAT AATCTTTTAA AAGCAGATT TAAAAAAT TTCAACGATT	37620
TTTTTTTAAAT GTTCATTGGT GTTTTGCCTG CATGTATATC TGTGTGAGGG TGTGAGTTT	37680
CCAGGAACTG GAGTTACAGA CAGATGTGAG CTACCTGTTG GTGCTAGGAA TTGAACCCAG	37740
GTCCTCTGGA AGAATAATCA GTGCTCTTAC CCACTGAGCC ATCTCTCCAA CCCAAATACA	37800
TCTTAAAAAA AATTAAAACA GTGGACCTGC CTTCTAGTTC TGTAGCATTG GCTACTCAA	37860
ATGACCCACA TCTTTCATCC ACCGTAGGCC TAAATCTCTT ACACTTATGA AACCTCTGA	37920
AAGAGGATGA GTAGACCTGA CCAGCCGAGA GACATCAGAG CCAGCAAGGT TAGGAAGCCT	37980
CTAGGAATGA CCCACACCAT AGAAATATTC TCTTTGCCCA GACACTGAAG TTATCTGAAA	38040
TGAGATACTG GTTATCTAGG ATGAACCCCG AGAGAAGAAA GGTTTAGGAC AGGTGTGGTG	38100
GTATGTGCCT TTAACCCAG CATTTAGGAG GCAATGGCAG ATGGACCTCT TGTGAGTCTG	38160
ATATACATTG TAAAGGGGAG AACTCCCGGA ATTTGTTCTC TGACCTACAC ATGTGACATG	38220
CATGTGTTTC TGCACACACA CATAACACACA CACACACTGT AAAAATGCAA AATGGCTACC	38280
AAGTGGTCAT TGAGCTTCTC AACCTCACTG ACAGCTACAT TATTATATAG ACTTACTGGG	38340
AACAGATCCG CAGGAAATTA TTTGGAATCT TTTTCTTTTC TCTAACGGGG GCTGATCTGG	38400
AACTTCTGAG CCTTTTTGTT CCCTATCATG AATGCTGGGA TGGCAGGCGT TTCCACATGA	38460

CTCGTTCGAT	G TAGTATTGC	AGACTGAACG	CAGGACTTTC	CACACACTAA	GCAGGCATTC	38520
TGTGAACTGT	TACGTCTCCA	GCCCCATTTT	TAAATTCTAA	CACCAAAGTG	CTAGTTTTGT	38580
CCCTTGACCT	GGACACTGCA	GTGAGTTCAC	AGAACTTATA	ATCACCTGT	TTAGTG TAGA	38640
AGCTACCTCA	ATCACCATGA	CATTTTTTCAA	AAATGTGTTC	ACTTTCCTCT	TTAGAGTCCA	38700
AGCACACCAA	GCTTGGCGGA	ACAATGATAC	AGTCTAACTG	GATCTGTTTC	AAAATTGCAA	38760
CTTGACTCTA	CATCTAAATA	GGTATGTGTT	GTGACAAGTT	TATTATGTTG	TGTGTGTGTG	38820
TACACATGTG	CCACAGGAAG	CCAAAGGACA	ACTTGCTAGA	GTGCATTTTC	TTTCCTGGGA	38880
ATTGGCCTCT	GGTTGTCAGG	CTTGGTAGCA	CGCACTTTGA	CCCTCTAAGC	CATCTTGATG	38940
GCCCAGAGAG	TGAACCACGC	TGTTTTCACT	TTCCTACTTC	TTGGGCTGAA	TTCTCAAGTA	39000
CCTGCCCTTG	CAGCTTTGCA	CCCTTCCTAA	CTTCAAAAGG	AAACTGACAT	GGAGAAGGGT	39060
GATACTTGAG	GATTTCTGG	CTCACTTAGC	TCAGGACTCT	GGCCTAAGAA	CAGGGAACCC	39120
AGCAGTGTGA	ACAGGGGTCC	AAGAGAGTTC	ATTTGTACTT	ACCGGCAAAA	CAGTGTGGCA	39180
GGCTTCACAC	AAATACATAC	TCGGCACCAG	GACAGGGCCA	CTCTGGATGG	AGGTGGGCTT	39240
AGGTGGGGTA	CTGCCCACCC	AGGGTTGTCC	TCTCTTGTA	GCAGACTCAT	GGGGACAGCC	39300
CAGAAGTGAT	CCCACAGTCT	CTCTGAAGCT	GACAATAGGG	GATAATTCTA	AGTCCTCATC	39360
CTGTGCTCAT	CCACAGTCCT	TTGTCGATCT	GGACACTACT	ATCATGGGCT	GCTGGAAACA	39420
GGTCTTTGCA	GCCCAAGTCT	GAGCCACTAG	CTCTGCTTTC	ACTGCCAGCC	ATTAACCTCC	39480
GGGAGTGGGC	GTGGGATAAG	AAGAAACATT	TATAGAGTCA	ACGGCCAATC	TGTATTTGGG	39540
CTGAAAACCA	TATTAAGGAA	GGGCCAAGCC	TGGCATAATG	GTGACCAGAG	CCACTAGGGG	39600
ACCAACTGCA	CCCAGCTTTA	GCAAAGTGAC	AGGCAGCATG	AGGTACCATT	ATGTGTGCTG	39660
GGCATGCGGC	TTCAGGATGG	CTCTGTGACC	TCCTAGAGGT	TGTCTTATTG	GCAGGCATAG	39720
GAAACAAAGG	CAGAGAATGA	ATGCTACAGC	CAGAGAGACC	CAGATCTGCT	AAGTGGATGA	39780
CTCTTG TACA	TATGTGTGTA	TGTTGTTTTT	GAGGCAGGGT	CTCACTGTGT	AGCTCTGACT	39840
GTCCTGGAAT	TGGATCTGTT	GGTCTCAAGT	TCAGATCCTA	GTGGTTTATT	TTTCCTGTGT	39900
ATGTGTGCTT	GTCATGCACA	AGCATGTGTT	AAGGTGAGTA	GATATGTAGG	CACATGGAGA	39960
TCAGAACAAT	GGTGTCACCT	CAAACCTTTA	TAGACCTATA	TCCATCTTGA	CATTAGGGTT	40020
ACAGGTGTGT	TCAACATAGA	TATGGCCAAA	ATTTAATGTG	GGTTCTGAAG	ATCTAAATAT	40080
GTCTTG TGTG	GGCTAGTTCT	ACGTCAACCT	GACACAAGCT	AGAGTTATCT	GAAGGAAGGG	40140
AACCTTAGTA	GAGAAACTGT	CTCCATGAGA	TCCAGCTGTA	TAGCATTTTC	TTAATTCCTA	40200
GTTAGAGACT	AATGGGGGAG	GGCCAGTCC	ATTGTGGGTG	ATGCAACCTT	AGACAGGTGA	40260
ACCTGGGTTT	TGTAAGAAAG	CAGGCTGAGC	AAGCCATGAG	GAAGCAAGCC	AGTAAGCAGC	40320

ACTGACCATG GCCTCTGCAT CAGCTCCTGC CTCCAGGTTC CTGCCCTGTT TGAGTTCTTG	40380
TCCTGACCTC CTTCCGTGAT GAACAGTGAT ATGGAAGTAT AACCAAATAA ACCCTCTCCT	40440
CCCAAGTTGC TTTGGTCATG GTGTTTCATC ACAGCAACAG AAAGCCCAAC TGAGGCAGGT	40500
TTTCATGCTG TATAACAAGC TGAACCATCT TACCAGCTCC ATAGTGTTTA TTTTAAAAGA	40560
TGAGTGTA ACTTTCCTTT TTTTCCTTTT AAAAATCCAA AGAACCACGT TCCTCAGGAA	40620
AAGCTCTGGG CCAGTTCTCC TGGTAACTTT GAAGTCTTTT TAAAGGCAGA GTCTATGTTA	40680
GACAAGCTGG CCTCAACCTC ACAGAGATCA CCTCCCTCTG CCCGTTAACT GCCTGGTGAG	40740
CTACAATGTG TTTTAAAGA TGTCCCTGTT CCCTCTTAAA CAACTCCAAT TTCACCCATG	40800
TGTTCCCATTT TGGTAGGACA GGAAGCCATT TGTTCATCAT GAAGCTTCTG CTGATGTCAG	40860
GACAGGCGCG CGCGGCACA CACACACACA CACACACAGC AGCTTTAGTC ATTTGTGGTC	40920
AGCTGGGAAA ATGGGAAAAC ACGGTTGGAG CTGAGTTGAA CTGAAGAGTT GGTGGAGACA	40980
CATGGTGCAA ATCCTGAGCA GTAGCTGAAG GAAAGGTACA AGTTTGGCAG TAGATTGGCC	41040
AATGAGGTGC AGAGATAAAG CAGAAGGGCT GCGGCGAGAG CTGCAGCATG GTGCGTGGA	41100
CCCTTCAGGA GGTAGAAAGG TAGAAAGGGC TGCTTGGACT ACTAGTGTGT AGATTACTGT	41160
CTTTCAGCAG GTGAAAGACA AGGCTAGAGC CTGTGATTGG ACAGTAGAAA AGGAGGGCGG	41220
GCTGAGAGTT TGAGAGTCTG GAGGGATAGG AGGAAAGAAG GAAGATGGAG GAAGAGAAGG	41280
ATGACCCAGA GCTGTGTGGC TTAAATAGC CACAGGTAGC TATGAATATC ATATAAGGGG	41340
TGGATTATGA CAGGACAATT TGTCCACTCA AGGTGGGCAG CTTATATCAT ATTAATTGGC	41400
TCTGAGTTCT TTGTCTTGGG CATTTTGTGA GCTGAGAATT TACTGATATA AATCTGACTG	41460
ATAAATTACA AGCCTCTAGA GTTTTGATTT TACTGGGTTA CAGGGATTTG TGACAGTTAA	41520
CTGCGAGATG CTACAGCCAG AGAGACACGG ATTCTGCTAA GTAGATGACT CTTGTACATA	41580
TGTGTGTATG GGGGCTAGCT GTGAAGGCAG TGAAACTGCT GCCAGGGCCA GAGAGTAGTT	41640
GGCACTACTG TGGGATGGTG CATCCATTTT TTTAAAATT ATTTAATGCA AACTAGTGA	41700
GTCATCCAGT AGGAAATGCT GGGGTCTGGG GAGCTGGGGG TGGAGGAAAG CCACAAGCCC	41760
ACGGAGCCCC AGATCCCCCA CCTCTTTGGA GAATAACACT GATATCAGTG ACTCAGACAC	41820
AATAGATCTT GGGGTTTCTG ACCCAAGCTC CTCTAGTAAG CATGGGTGCA AAAGGTGTGG	41880
AATGGAGAGT GAAGGAAGAC TTTTTCATAA GCCTGTCACA AATGAGGAGG AAGCTAAGCT	41940
TGGGAAATGC AGGCCTTCAG TGGCAGACCA AGTGGAGTCA ATGAAGTAAG GTCTGAGTAG	42000
AAGGGCTCTG GGTGTGCGCT TCAGGCTGGG TGCACACTTC TTTCTGAGGA AATGCTCACT	42060
TCCACTTTGA CCATTCCCTG ACCCAGGTCA TAGCTGATGT GCCAGAGTGT CATGGGTGAA	42120

GTGGTCACTT	TCGCTCTTTC	CACACAACCTG	TGCTGTGTAA	GACACCCTTT	CTCTGGTCAT	42180
ACAGGAGTCC	CCTGTGGGGT	TTGAGCCCTG	ACTTAAAAAG	AAAGGATGAG	GGCTACTTCT	42240
GTGGAAGGGA	GCAAAGAGCA	GAGGTCATTG	CTGCTGGAGG	AGATCTGCTA	ACAAGCATGT	42300
GATGTTTAAC	ATTAAGGGCT	GCTCATCAAG	TCAGCACTGA	CTCCAGCAGA	GTCCTGTCTA	42360
GGCTACTCCA	GTATGCCCTG	GTCAAGACTA	GCCTTGCCAA	GGGAGCAGCC	TGGGCTGTTG	42420
CTAGGTGGAT	AGAAGCACAC	ACAGAGGATT	TTCTTAGTGG	TGTATGTAAT	CACTGAGGTC	42480
TTGCTGACCC	AGTAGGCATA	CTCCTCCATT	GCTAGACTCA	GTCACACAAA	GTGTACAAGA	42540
ACAGGGCATT	CTTCATGGAA	AATTCCTGAC	TGGGTCTTTT	AGAGCTCCAG	TTCCTAGAGG	42600
GGCAGATGAT	CCCAGTGACT	TATGCTCAGT	GTAAAGCTGG	TCTGCTGTCA	CATCTTTGCT	42660
CCCAAAGGTT	TCTGGGATTC	CTCCTGTACT	TTCTTCTATT	TTTATTTTCA	AGACAGGGCT	42720
TCTCTGTGTA	GTCCTGGCTG	CTCTGGAACT	TGCTCTGCAG	ACCAGGTTGG	CCTCAAATTT	42780
ATAGAGATCC	ACTTGCCTCT	GCCGTCCAAG	TACAGGGATT	AAAGTTGCAT	GCCACCACTG	42840
CCCAGCCTCT	CTAAAATTTT	CTTAATTAAT	TTATTTTTC	AGACAGAGTC	TCATATGTA	42900
GTCCTGGATA	TGCTGGAACT	CAGTAATGTA	GAACAGGCTG	TCCTTGAACA	TACAGAGTTC	42960
CACCAACCTC	TGCTTCCAAG	TGCTGGGATT	GAAGTGTGTG	CCACTATGCC	CAGCTAAAAC	43020
CTGTTTTATT	TTCTGTGCAT	GGGTGTTTGC	CTGCATGTAT	GTCTGTGCAT	CATTTGCCTG	43080
ACTGGTGCCC	ACGGAAGTCA	GAGGAGGACA	CTGGATCCCC	TGAGGTGCCC	ACGGAAGTCA	43140
GAGGAGGACA	CCGGATCCCC	TGCAGTGCCC	ATGGAAGTCA	GAGGAGGACA	CCGGATCTCC	43200
TGGATCTGGA	TGACTGAGCC	ATCACATGGG	TCTTGGGAAA	AGATCCCGGG	TTTGCTCTAA	43260
GAACAAGTGC	TCTTAATGAT	TGAGATGTCT	CTCTATCCCA	TGTTTCTTTG	TACACAAACA	43320
CCATGGACAC	GTGGCATAAC	CTGGGCTTCC	TTTTACACAC	ACTCTGTCGA	ACTTAAATTC	43380
TGCTGGCGGC	TCCAACCTGAC	CTTTCCTTTC	TATTCCTAAA	TTCTCGGCAT	GGCTTGGGTC	43440
TGGTTAAGTC	CCCCCTTTTC	CAAGCAGCCG	GAAGCACTTA	TCTCTGAATG	TGCCTCTGTG	43500
GGACACACCG	GGGGACCTGC	TGAAGCCTCT	GAAGAGCAGA	GGTGATGTCT	GCCTCCCCAT	43560
CTTTGCCCTC	TTGTGCTAAG	AAGCTACTTG	TGATGCTGGA	GGTGGTGGGG	AAAACCCACC	43620
AGCCTTGCCA	CCTGAAGTGA	AGGGCAGCCA	CGGCCTGTGT	CCTAGCCAGT	GGGGATTAGT	43680
GAAAATGGTA	AAGTGGGCAA	CGAGGCTGCT	TGCTTTCTGA	GCTTCCTCCT	ATTTTGGGTT	43740
GGTAGCAGCA	GCGGCCAGT	TCCTTCCCAC	TGTGGGGATG	AGGAGTACGC	CCTCAGGATG	43800
CCGGCATCAG	AGAAGGCAAG	AACAGACGCA	GTGTCGCACG	TCTTCAATTA	CAGCACTTGG	43860
GAGGCAGAGA	CAGGCAGATC	TCTGCGAGTT	CAAGGCCAGT	CTGGTCTACA	CAGTGAGTTC	43920
TAGGTTGCTC	TGTGTTACAC	AGGGAGAACT	GTCTGAAGAA	ACAAACAAAG	AGAAAATTAA	43980

AGTTAGATGT AGTGGCACGG TCATAATCTA AAATGTGGCC TAGCTGTTCT CTGTTCTCTG	44040
TTTCTCTTTC TTCCTCCCTC CCTCTCTCTT CTTCAATTGTC TGTCTGTCTG GTGCTTGTAT	44100
ATCAAAATGT AAGTTCTAAG ATATGCTTCA GCACCGTGCC TGCCTGCCTG CCGCCATGCT	44160
CCACCATGAT AGTCATAGAC CCACCCTCTC GAACTGTGAA TCCCAAATTT ACTTTCTTCT	44220
ATGAGTTGCC CTGGTTATGG TGCCTTATCA CAGCAACAGA GCAGTGAGTA ATATACCCAC	44280
CCTCAAAGAC AAGCTGAAAG AGAGACCCAT GTGCTGTGGC ATGCGTGTGC CTACACTTAA	44340
CACACATAAA TAAATACATC TCCTGAAGAA AATTTAAAAG TTATTCTGGA CAGAACTAG	44400
AGAGGCCAGA CTGGCCTCAG CTCAAGCCCA CAGCAGCTCC TCTGTCCTGC TGTCTTTTCC	44460
TGTAGAGAAA TTCAGTGAGA CCAAGCTGT CTGTCCTAGG GCTATAAGCT GGGTGGGTGG	44520
CTGGGATGAC CAACTTGAT AGAAAAGAGG AAAAGGAACT GGGAGTTGCG GCCGCC	44576

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGACAGCCCG AAGGACTACA GGT

23

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

41

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CGAAGAACTC CGCAGGGTCC

20

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

AAGACCCGCC ACGACCCG

18

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GAATCAGCAC CCTCTCCGCC

20

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

42

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

TGCGGAGTTC TTCGTGCTGA TGGAG

25

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGTGCTCGGC GCGTCCTTC

20

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

GAGTGGCGGA GAGGGTGCTG A

21

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

43

- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GGCCGAGGCT GAGCGGGG

18

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CTGAAGGACG CCGCCGAGCA

20

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Synthetic Primer"
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

CTCCAACGCC TGCCGCTGC

19

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:

44

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic Primer"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GCAGGAGGAG CGGGAGCAGG A

21

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "Synthetic Primer"

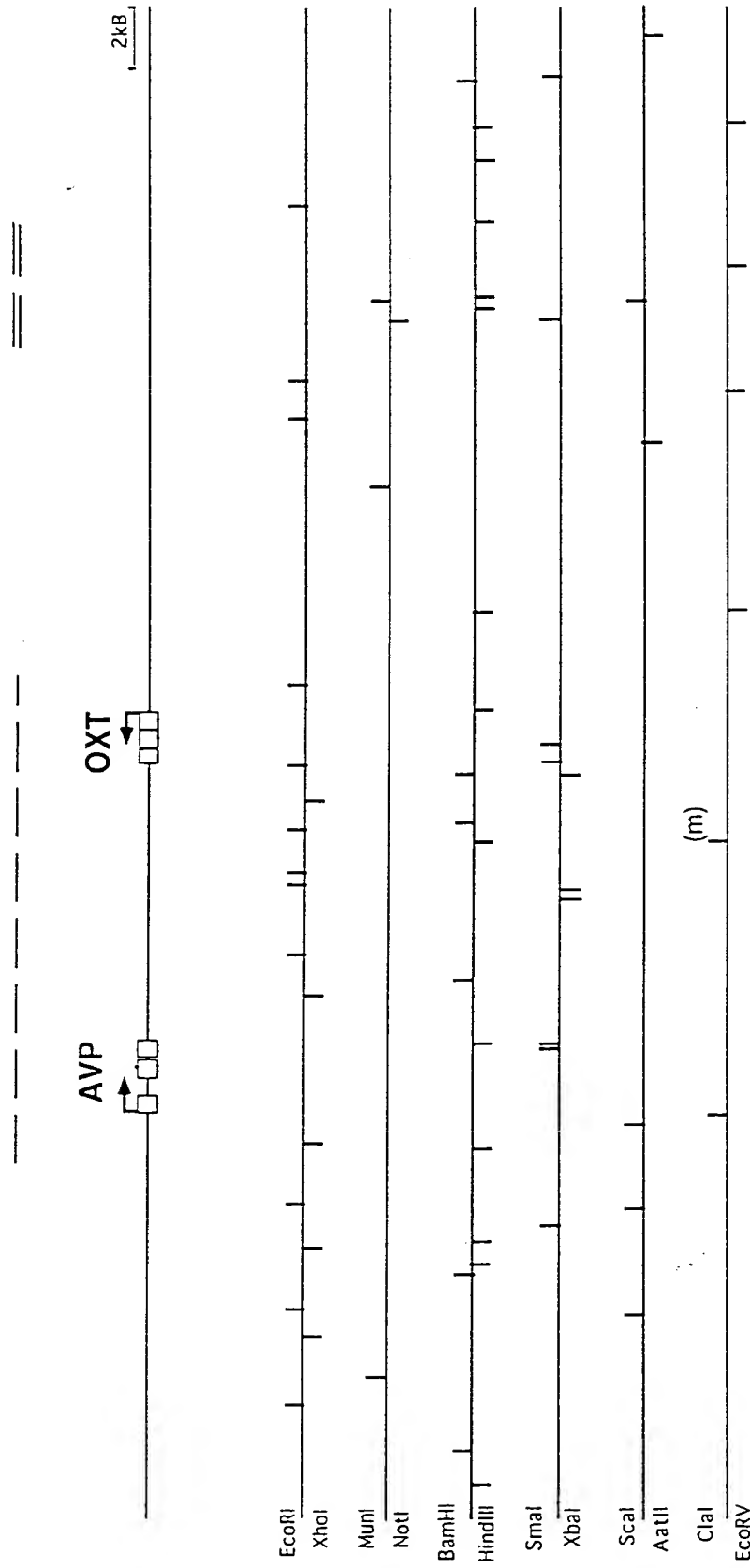
(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

TCCAGTGCCC CGCAAGCCG

19



(m) denotes methylation

FIG. 1
1/14

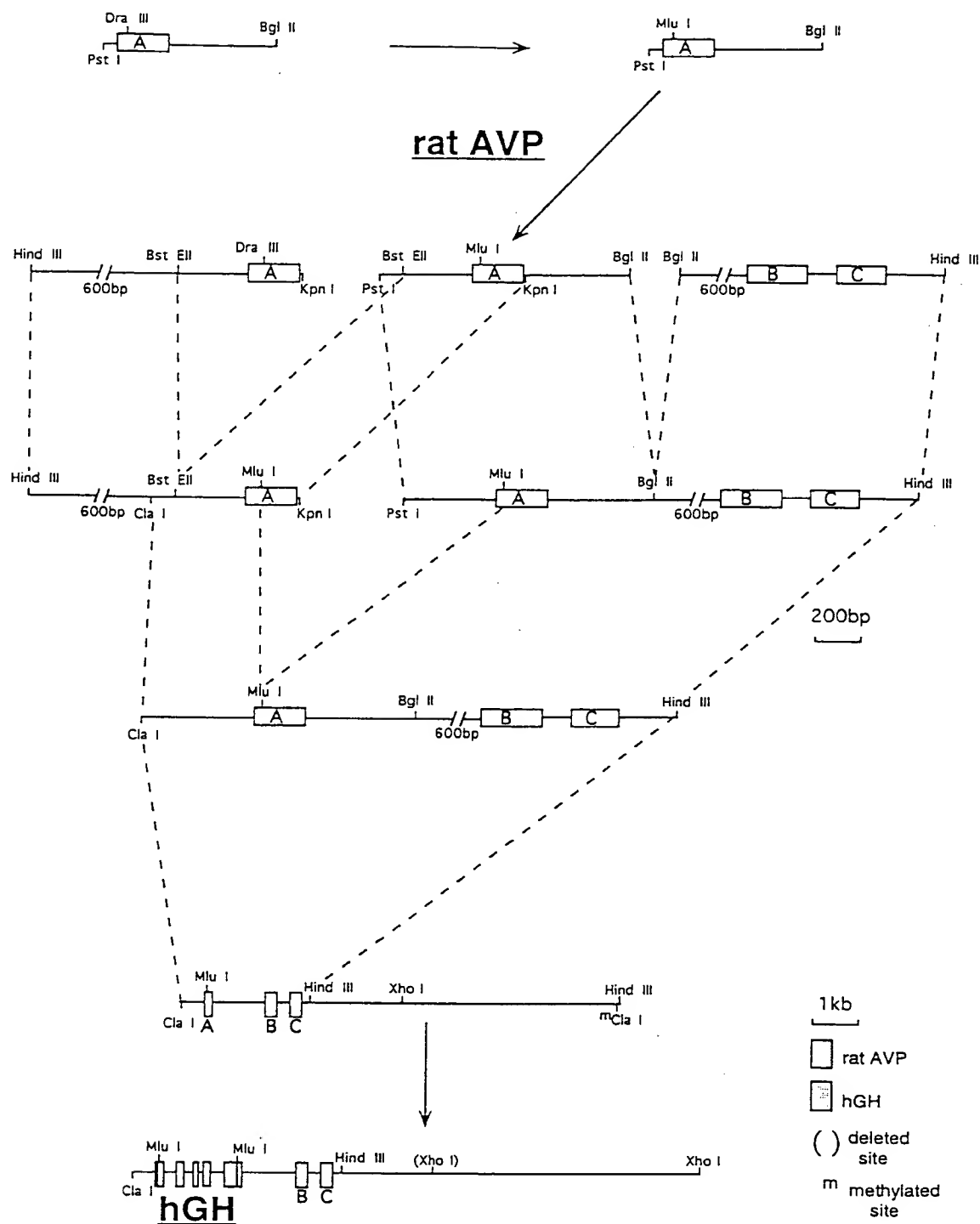


FIG. 2
2/14

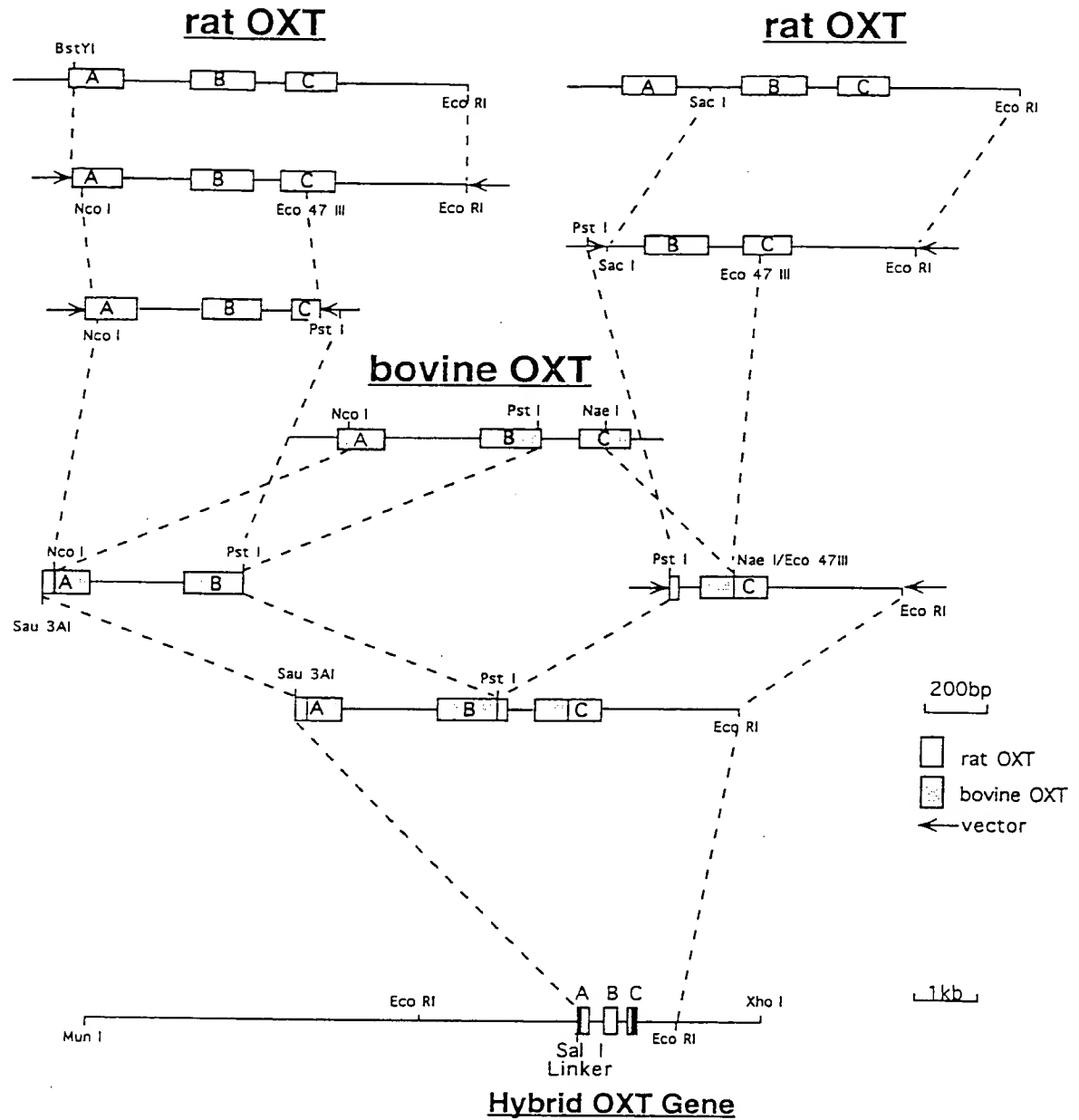


FIG. 3
3/14

rat AVP/OXT gene locus

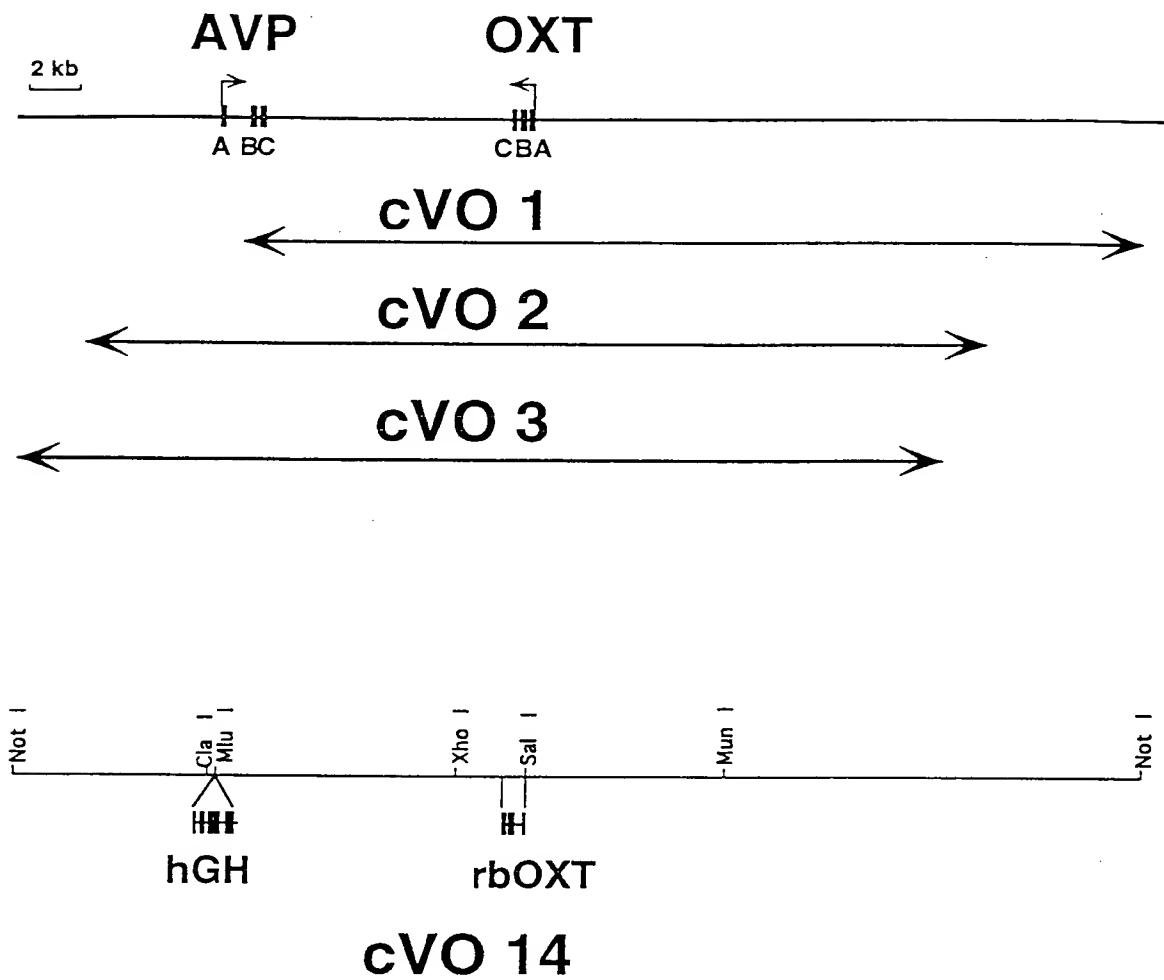


FIG. 4

4/14

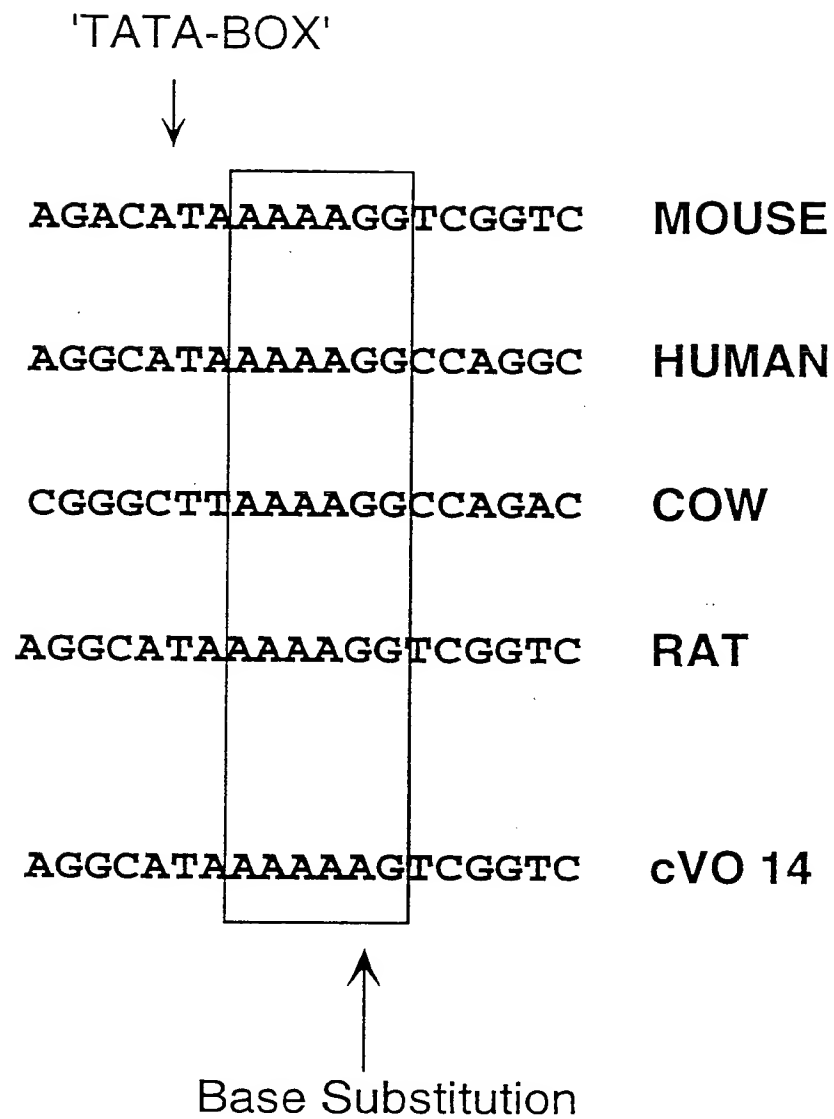


FIG. 5

Figure 6 5'OT-EST PROTEIN OF DIFFERENT SPECIES

Mouse

MLRALNRLAQRPGDRPPTPLLLPVRGRKTRHDPPAKSKVGRVQTPPAVDPAEFFVLTERY
 GQYRETVRALRLEFTLDVRRKLHEARAGVLAERKAQQAITEHRELMAWNRRDENRRMQELR
 IARLQLEAQAQEVQKAEAQQRQRAQEEQAWVQLKEQEVCLKQEEAKNFITRENLEARIEEA
 LDSPKSYNWAVTKEGQVVRN

Rat

MLRALNRLAARPGGQPPTLLLLPVRGRKTRHDPPAKSKVGRVKMPPAVDPAELFVLTERY
 RQYRETVRALRLEFTLEVRGKLHEARAGVLAERKAQEAIREHQELMAWNREENRRRLQELR
 IARLQLEAQAQELRQAEVQAQRAQEEQAWVQLKEQEVCLKQEEAKNFITRENLEARIEEA
 LDSPKSYNWAVTKEGQVVRN

Human

MLRALSRLGAGTPCRPRAPLVLPARGRKTRHDPLAKSKIERNMPPAVDPAEFFVLTERY
 QHYRQTVRALRMEFVSEVQRKVHEARAGVLAERKALKDAAEHRELMAWNQAENRRRLHELRL
 IARLRQEEREQEQRQALEQARKAEVQAWAQQRKEREVLQLQEEVKNFITRENLEARVEEA
 LDSRKNYNWAITREGLVVRPQRRDS

Alignment

Mouse	MLRALNRLAQRPGDRPPTPLLLPVRGRKTRHDPPAKSKVGRVQTPPAVDPAEFFVLTERY
Rat	MLRALNRLAARPGGQPPTLLLLPVRGRKTRHDPPAKSKVGRVKMPPAVDPAELFVLTERY
Human	MLRALSRLGAGTPCRPRAPLVLPARGRKTRHDPLAKSKIERNMPPAVDPAEFFVLTERY
Mouse	GQYRETVRALRLEFTLDVRRKLHEARAGVLAERKAQQAITEHRELMAWNRRDENRRMQELR
Rat	RQYRETVRALRLEFTLEVRGKLHEARAGVLAERKAQEAIREHQELMAWNREENRRRLQELR
Human	QHYRQTVRALRMEFVSEVQRKVHEARAGVLAERKALKDAAEHRELMAWNQAENRRRLHELRL
Mouse	IARLQLEAQAQEVQKAEAQQRQRAQEEQAWVQLKEQEVCLKQEEAKNFITRENLEARIEEA
Rat	IARLQLEAQAQELRQAEVQAQRAQEEQAWVQLKEQEVCLKQEEAKNFITRENLEARIEEA
Human	IARLRQEEREQEQRQALEQARKAEVQAWAQQRKEREVLQLQEEVKNFITRENLEARVEEA
Mouse	LDSPKSYNWAVTKEGQVVRN
Rat	LDSPKSYNWAVTKEGQVVRN
Human	LDSRKNYNWAITREGLVVRPQRRDS

Predicted deleted form in JP17

MLRALNRLAARPGGQPPTLLLLPVRGprprsfssapfssqds

↑

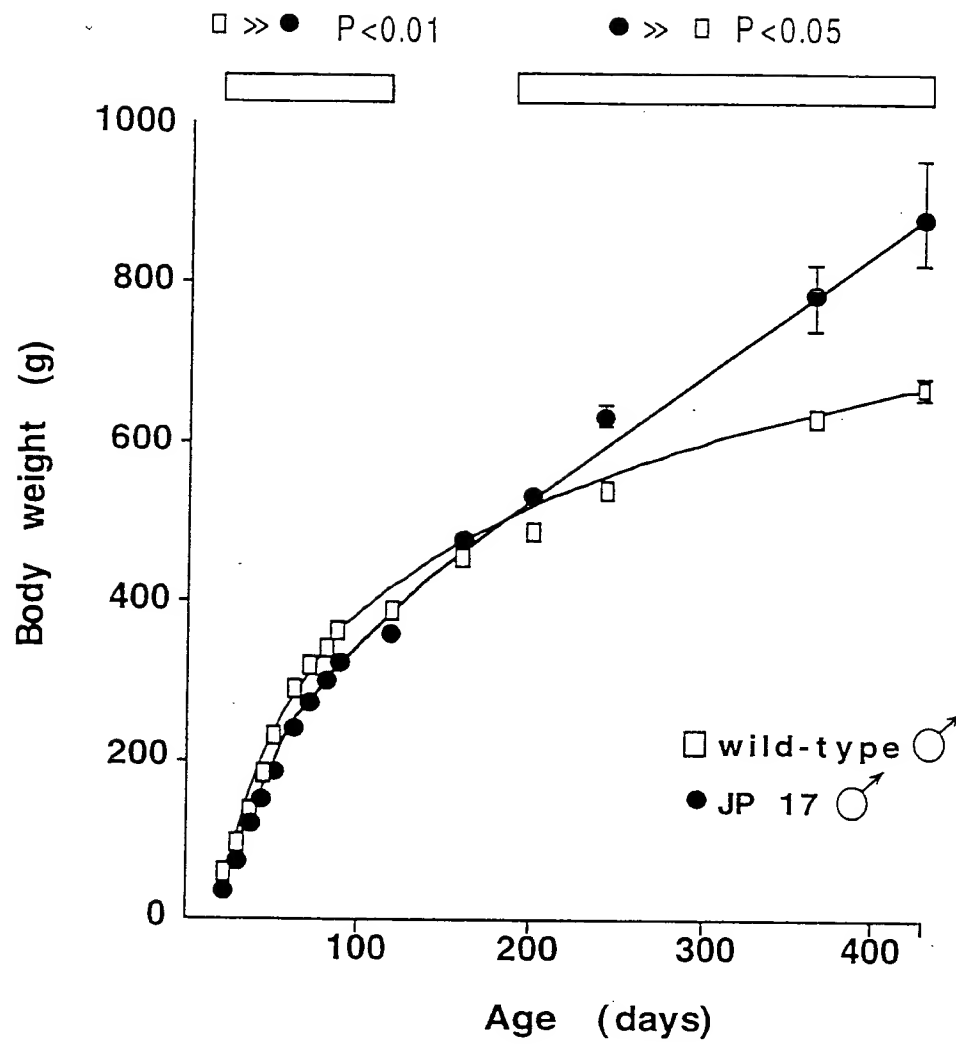


FIG. 7

7/14

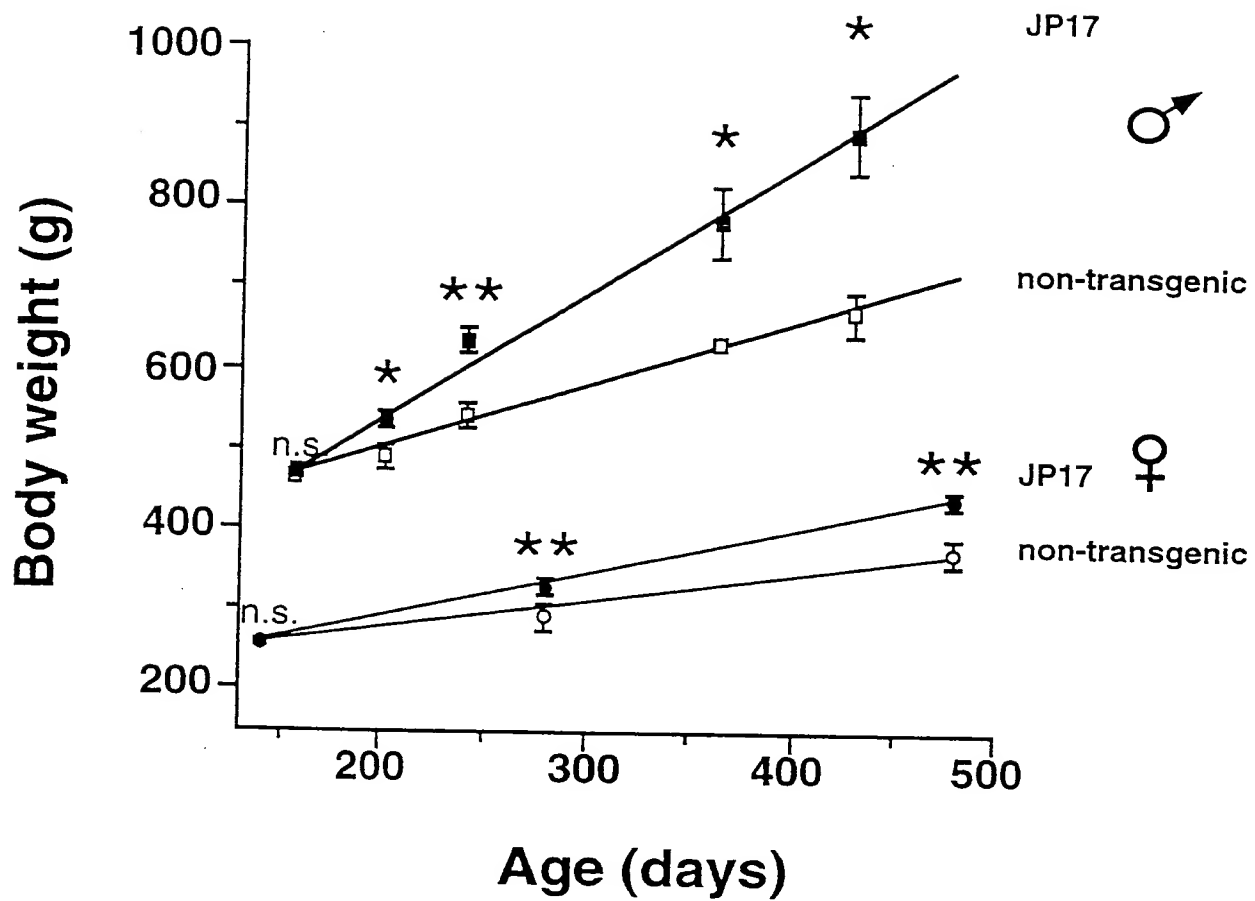
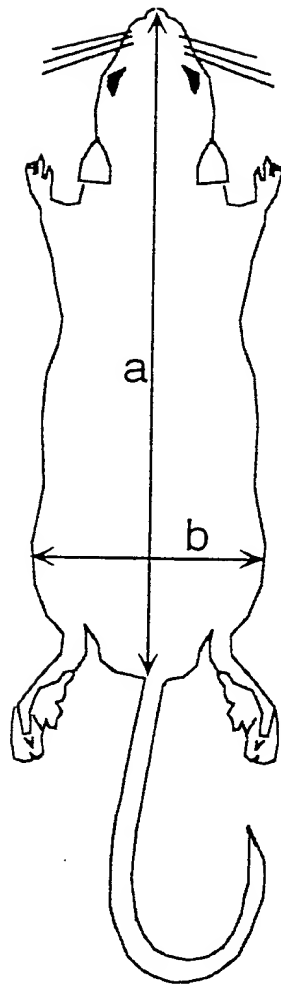


FIG. 8

8/14

20 weeks

	non-transgenic	JP17 transgenic
a	260.33 ± 0.28	243.34 ± 0.13 ***
b	92.67 ± 0.29	115.14 ± 0.24 ***

52 weeks

	non-transgenic	JP17 transgenic
a	273.83 ± 0.28	261.0 ± 0.45 ns
b	113.83 ± 0.10	157.83 ± 0.61 ***

FIG. 9**9/14**

$\frac{\text{body weight}}{\text{body length}}$
(g/cm)

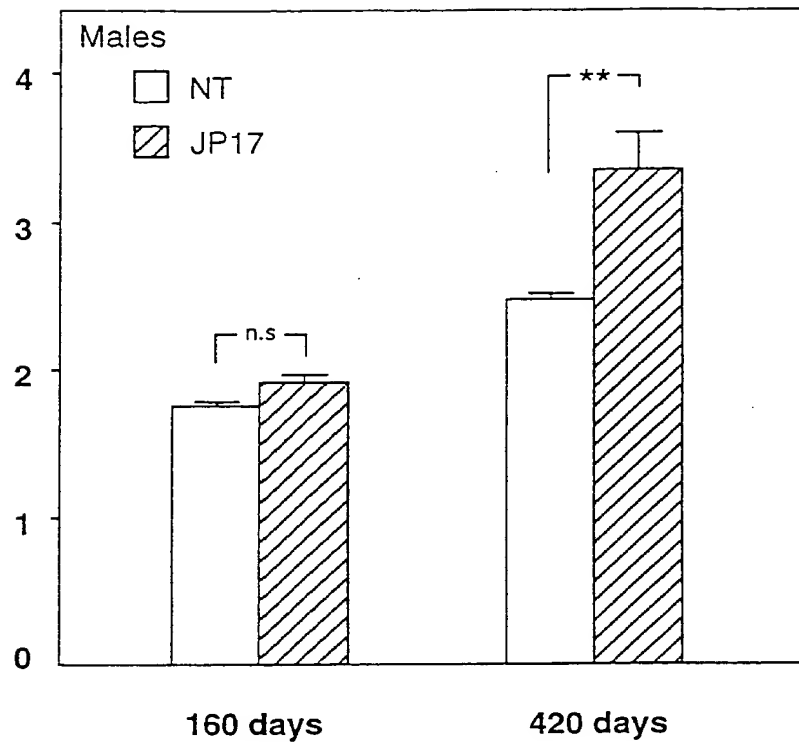


FIG. 10

10/14

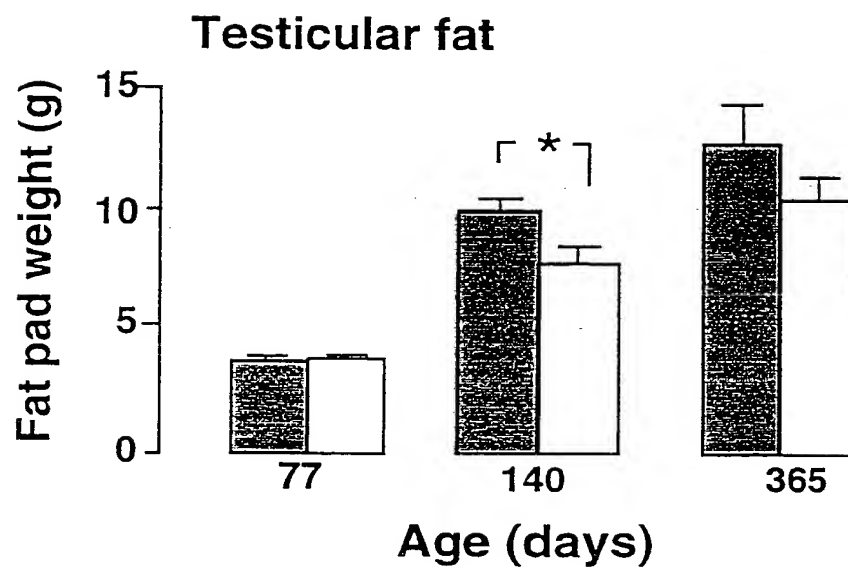
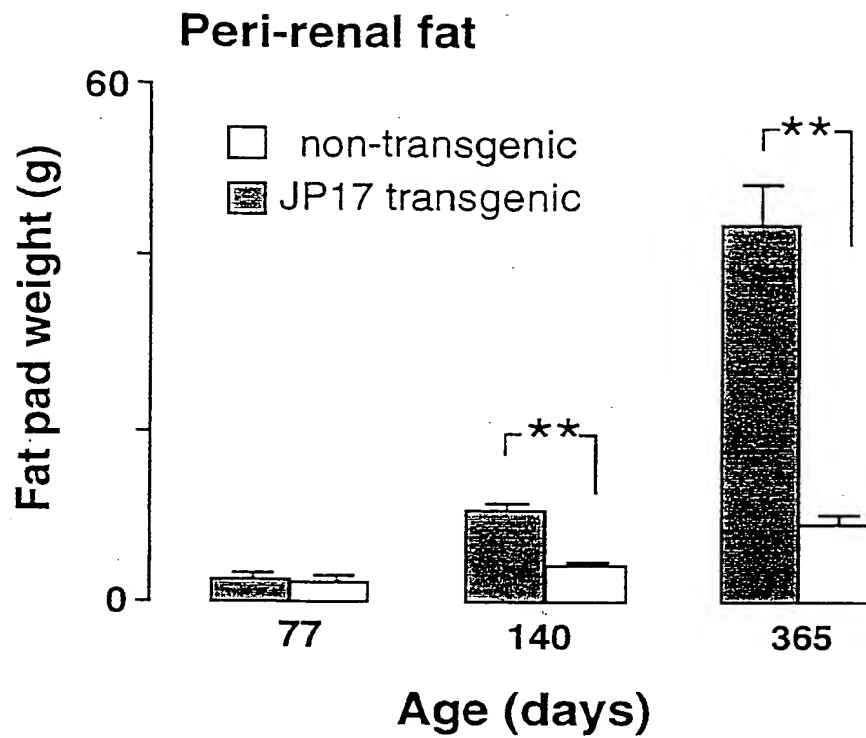
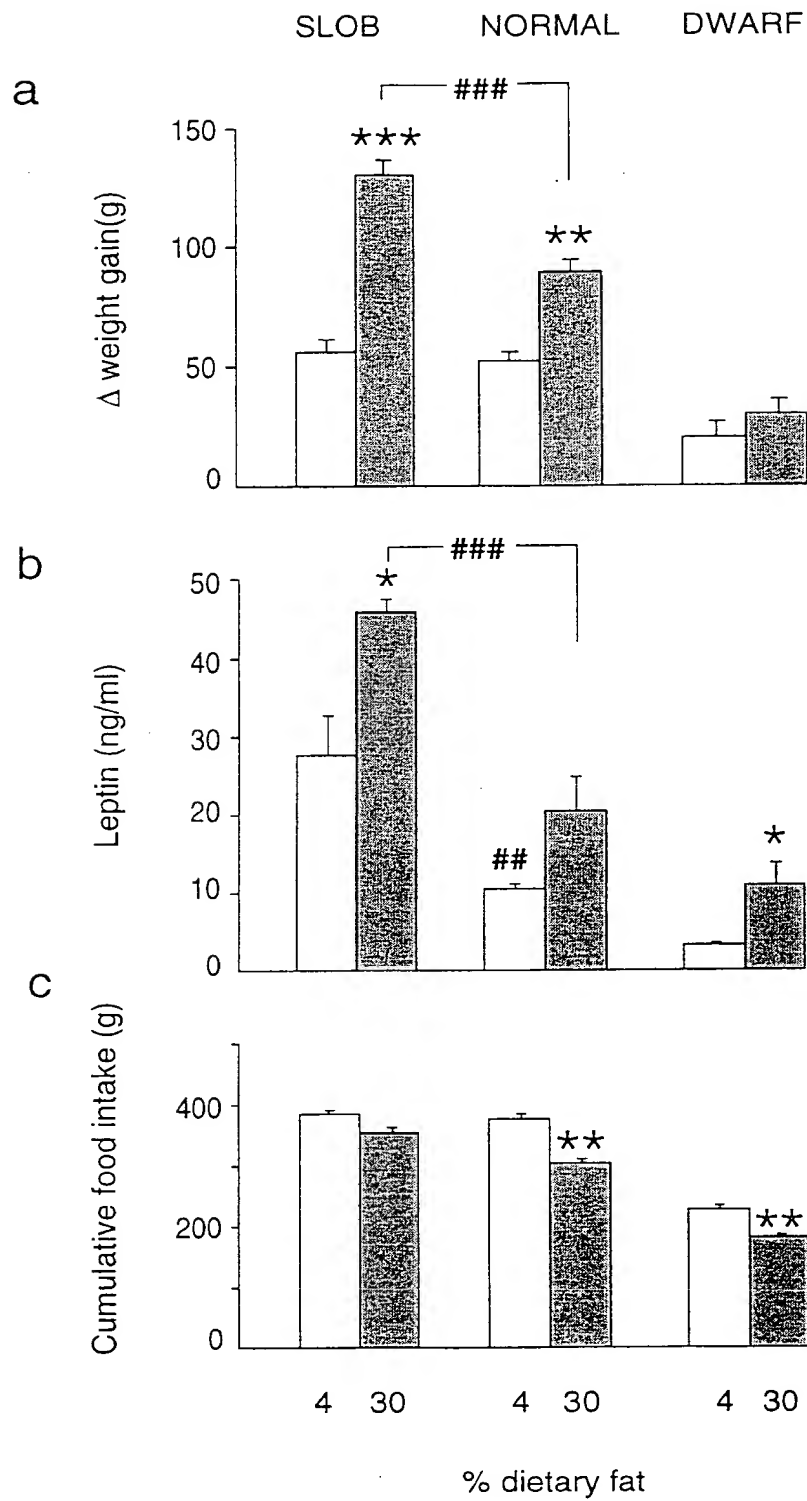


FIG. 11

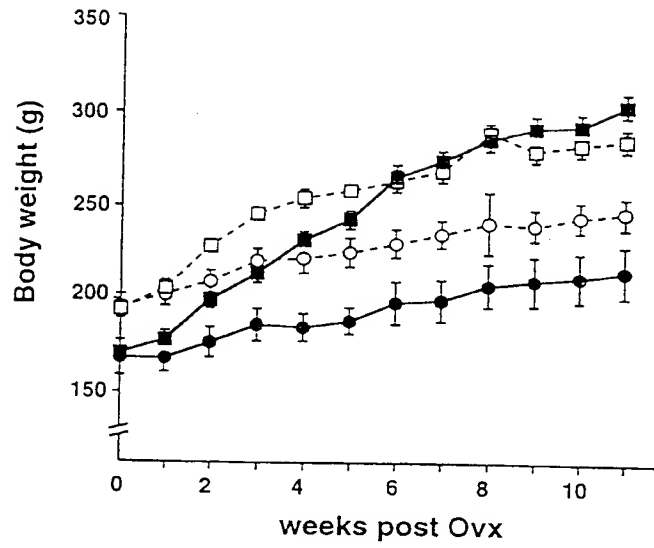
	Cholesterol mg/dl	Triglyceride mg/dl	Glucose mg/dl	Insulin ng/ml	Leptin ng/ml	Corticosterone ng/ml
MALE TRANSGENIC	122.3 +/- 6.4	*295.6 +/- 28.7	114.7 +/- 4.2	1.94 +/- 0.89	*24.4 +/- 1.49	168.9 +/- 23.5
MALE NON- TRANSGENIC	129.9 +/- 9.3	178.9 +/- 23.5	121.0 +/- 3.9	2.8 +/- 1.93	9.51 +/- 2.14	113.9 +/- 20.3
FEMALE TRANSGENIC	94.9 +/- 5.9	224.2 +/- 52.3	126.3 +/- 3.3	2.51 +/- 0.64	*14.74 +/- 1.38	256.3 +/- 104.1
FEMALE NON- TRANSGENIC	100.2 +/- 8.0	195.5 +/- 34.5	135.4 +/- 6.7	2.54 +/- 2.32	4.58 +/- 0.47	349.3 +/- 123.7

FIG. 12

12/14

**FIG. 13**

13/14

**FIG 14****14/14**